Functional analysis and structural investigations of *MTB* DosS sensory domain

Sunita Sardiwal

Department of Biochemistry and Molecular Biology
University College London

A thesis submitted for the degree of Doctor of Philosophy
July 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Statement

The work presented in this thesis is part of a collaborative project with Professor Neil Stoker at the Royal Veterinary College, University of London. Ultraviolet and Visible spectra were performed in the laboratory of Professor Peter Rich at University College London and Electron Microscopy conducted with the help of Professor Helen Sabail at Birkbeck, University of London. Except for the collection and processing of all NMR spectra by Dr Richard Harris I, Sunita Sardiwal confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been identified.

Part of this work has been published:

Sardiwal, S., Kendall, S.L., Movahedzadeh, F., Rison, S.G., Stoker, N.G., Djordjevic, S. (2005). A GAF domain in the Hypoxia/NO-inducible *Mycobacterium tuberculosis* DosS Protein Binds Haem. J. Mol. Biol. *353*, 929-936.

Abstract

Mycobacterium tuberculosis (*MTB*) is a very successful pathogen, causing the deaths of approximately two million people a year world-wide. Survival of the pathogen *in vivo* is dependant on its ability to respond and adapt to changes within its environment. One method of adaptation is through two-component signal transduction systems, the phosphotransfer pathways that couple stimuli to responses. Generic two-component systems involve two conserved elements, a membrane bound histidine kinase, which is the sensory protein, and a response-regulator protein that controls the response usually by altering the expression of genes required for adaptive responses.

The aim of this thesis is to investigate the structure and function of the sensory protein DosS (DevS). DosS is induced by exposure to hypoxia, NO and ethanol and is the only one of the 11 paired *MTB* two component systems for which inducers have been identified but the precise chemical nature of the signal is unknown. The N-terminal input region of the DosS sensor contains two putative GAF domains. Various fragments of the N-terminal region were cloned, expressed and purified to homogeneity.

Ultraviolet-Visible spectral analysis reveals that full length DosS binds a classical high spin b haem cofactor. Mutagenesis identified histidine 149 of DosS, which is within the N-terminal GAF domain, as critical to haem-binding. This is the first known GAF domain to bind haem and the presence of a haem co-factor is consistent with the postulated involvement of DosS in oxygen and redox sensing. Based on this data a model for histidine kinase activation is suggested. The second GAF domain of DosS was analysed using NMR spectroscopy. Triple resonance NMR experiments enabled the identification and sequential assignment for 99 out of the 141 backbone amide proton and nitrogen resonances. In addition, both of the GAF domains were also subjected to protein crystallisation while the full length DosS was investigated using electron microscopy.

Acknowledgements

This thesis could not have been possible without the encouragement and support of my family (my mum, brother and sister), friends and colleagues.

Firstly, I would like to start by thanking my supervisor, Dr Snezana Djordjevic for giving me the opportunity of doing a PhD and for her continual guidance and support throughout the course of this work. I would also like to thank all the other people in the lab who have helped me over my PhD years, especially Dr Mark Jeeves, Dr Syeed Hussain, and Paul Leonard. Thanks, goes to all of the people on the 6th floor of the Darwin building who have had made it a fantastic place to work.

I am indebted to Dr Richard Harris for his support, stimulating discussions and constructive criticisms throughout the preparation of this thesis.

I must also thank my collaborator Professor Neil Stoker and all the people within his lab at the Royal Veterinary College.

Finally, special thanks to Ashoke Chaudhury for all of his love and support, particularly in the last few months. This thesis is dedicated to him.

CHATPER TWO: MATERIALS AND METHODS

CHAPTER FIVE: STRUCTURAL INVESTIGATIONS OF GAFB

List of Figures

CHAPER ONE: INTRODUCTION

CHAPTER THREE: CLONING, EXPRESSION AND PURIFICATION OF THREE *MTB* SENSORY PROTEINS

CHAPTER FOUR: LIGAND BINDING; IMPLICATIONS FOR DosS FUNCTION

CHAPTER FIVE: STRUCTURAL INVESTIGATIONS OF GAFB

CHAPTER SIX: GENERAL CONCLUSIONS AND DISCUSSION

List of Tables

List of Abbreviations

ABBREVIATION	DEFINITION
NMR	Nuclear Magnetic Resonance
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
BCG	Bacilli Calmette-Guérin
BSA	Bovine serum albumin
cAMP	Cyclic Adenosine monophosphate
CD	Circular dichroism
cGMP	Cyclic guanidine monophosphate
CSI	Chemical Shift Index
cya B2	Cyanobacteria adenylyl cyclase
DosR	Dormancy survival regulator
DosS	Dormancy survival sensor
DrCBD	Deinococcus radiodurans chromophore binding domain
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylene-diamine-tetraacetic acid
EM	Electron microscopy
ESI-MS	Electrospray ionisation mass spectroscopy
GAF	Cyclic GMP, adenylyl cyclase, FhlA
HAMP	Histidine kinases, adenylyl cyclases, methyl binding proteins and phosphatases
HATPase	ATP binding domain
HisKA	Histidine kinase dimerisation
HK	Histidine kinase
HSQC	Heteronuclear single quantum correlation
IPTG	Isopropyl-β-D-thiogalactopyranoside
kD	kilo Daltons
LAM	Lipoarabinomannan
LB	Luria-Bertani

ABBREVIATION	DEFINITION
MTB	*Mycobacterium tuberculosis bacillus*
NEB	New England Biolabs
Ni-NTA	Nickel-charged nitriloacetic acid
PAS	Period, aryl-hydrocarbon receptor nuclear translocator and single-minded
PBP	pheromone-binding protein
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDE2A	Cyclic nucleotide phosphodiesterase 2A
PDE5	Phosphodiesterase 5
PDEs	Phosphodiesterases
Pfam	Protein families database of alignments and hidden Markov models
ppm	Parts per million
PVDF	Polyvinylidene Difluoride
RF	Radiofrequency
rmp	Revolutions per minute
RR	Response-regulator
SCID	Severe combined immunodeficiency
SDS-PAGE	Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis
SEC	Size exclusion chromatography
SignalP	Signal peptide
SMART	Simple modular architecture research tool
T7-pol	T7 RNA polymerase
TB	Tuberculosis
TEV	Tobacco etch virus
TFE	2,2,2 –trifluoroethanol
TM	Transmembrane
TM-HMM	Transmembrane helices based on a hidden Markov model
UV	Ultra Violet
UV-VIS	UltraViolet and Visible

CHAPTER ONE

INTRODUCTION

1. Mycobacterium Tuberculosis

Tuberculosis (TB) has been a global burden for centuries and brings about more deaths worldwide than any other infectious disease. Throughout history TB has earned nicknames such as The Great White Plague and The Captain of the Men of Death and even now it is estimated that approximately 2-3 million people die each year from this disease worldwide (World Health Organisation, 2001). The causative agent of TB was identified as *Mycobacterium tuberculosis bacillus* (*MTB*) by Robert Koch in 1882.

MTB belongs to the actinomycete branch of Gram positive bacteria whose members are aerobic, non-motile, non-spore forming, slightly curved or rod-shaped bacteria with an average size of 1.0-10 X 0.2-0.6 μm. *MTB* can be visualised using the Ziehl-Neelsen staining method, as their cell walls are acid-alcohol fast and therefore retain the red carbol fuchsin dye after decolourisation with mineral acids or alcohol.

The *MTB* virulent H37Rv strain was first isolated in 1905 and is the most commonly used strain in tuberculosis research. The complete genomic sequence of the H37Rv strain was first determined and annotated in 1998 (Cole *et al.*, 1998) but has subsequently been re-annotated by Camus and co-workers (Camus *et al.*, 2002). The genome contains 4.4 Mb of DNA (similar to the size of *E. coli*) with 4,006 protein coding genes and has a relatively high average GC content of 65.6%. Deciphering the sequence of the *MTB* gene has allowed for significant advances in understanding the biology of this bacillus and hopefully, in parallel with developments in scientific tools, will pave the way to eradicating TB.

1.1 *MTB* infection and disease

TB can develop anywhere in the body but usually presents as a pulmonary infection. Pulmonary infection can be divided into a number of key stages: inhalation, adherence to and ingestion by macrophages, replication within macrophages, T-cell response, extracellular growth in liquefied caseous lesions and transmission. Figure 1-1 illustrates the process of active *MTB* infection, however it is important to note that infection with the tubercule bacillius does not necessarily lead to disease; the bacilli may remain dormant within human tissues for decades without replicating (section 1.3).

Primary infection results from the inhalation of a droplet containing the bacillus which has been coughed up by an individual with active TB (Figure 1-1(A)). The droplet becomes lodged in the alveoli and the bacilli are engulfed by non-specifically activated alveolar macrophages (Figure 1-1(B)). Binding and phagocytosis of mycobacteria occurs

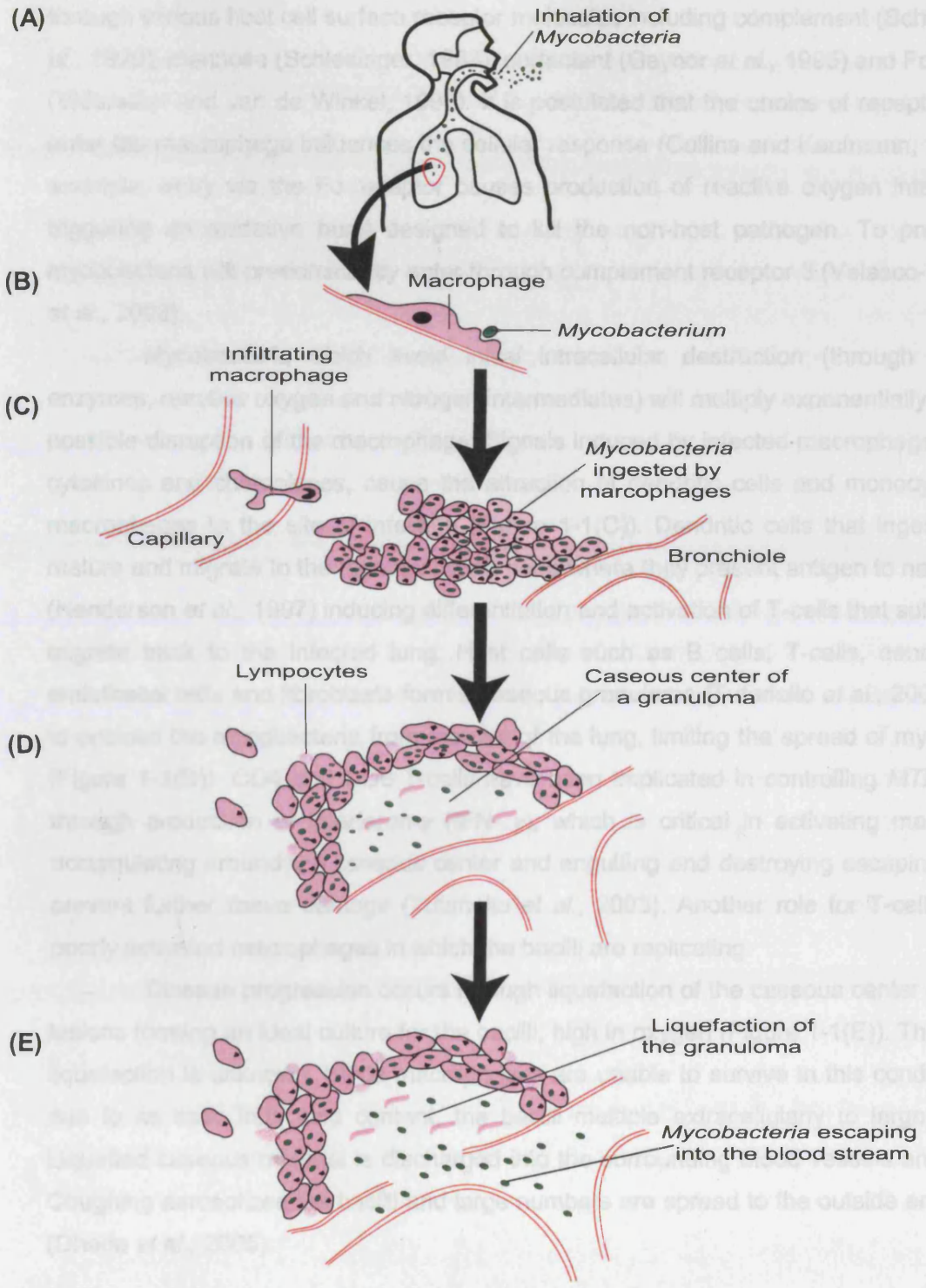


Figure 1-1: The stages of active pulmonary tuberculosis: A) inhalation of *MTB*; B) an alveolar macrophage ingesting a bacillus; C) macrophages migrating from the blood stream and replication of *MTB* within macrophages; D) formation of a granuloma; E) liquefaction of the granuloma and transmition of *MTB* into the blood steam. This figure was adapted and redrawn from (Russell, 2001; Schlossberg, 2006).

through various host cell surface receptor molecules including complement (Schlesinger *et al.*, 1990), mannose (Schlesinger, 1993), surfactant (Gaynor *et al.*, 1995) and Fc receptors (Vidarsson and van de Winkel, 1998). It is postulated that the choice of receptor used to enter the macrophage influences the cellular response (Collins and Kaufmann, 2001). For example, entry via the Fc receptor causes production of reactive oxygen intermediates triggering an oxidative burst designed to kill the non-host pathogen. To prevent this, mycobacteria will predominantly enter through complement receptor 3 (Velasco-Velazquez *et al.*, 2003).

Mycobacteria which avoid initial intracellular destruction (through lysosomal enzymes, reactive oxygen and nitrogen intermediates) will multiply exponentially leading to possible disruption of the macrophage. Signals induced by infected macrophages such as cytokines and chemokines, cause the attraction of dendritic cells and monocyte-derived macrophages to the site of infection (Figure 1-1(C)). Dendritic cells that ingest bacteria mature and migrate to the regional lymph node where they present antigen to naïve T-cells (Henderson *et al.*, 1997) inducing differentiation and activation of T-cells that subsequently migrate back to the infected lung. Host cells such as B cells, T-cells, dendritic cells, endothelial cells and fibroblasts form a caseous granuloma (Tufariello *et al.*, 2003) serving to enclose the mycobacteria from the rest of the lung, limiting the spread of mycobacteria (Figure 1-1(D)). CD4 and CD8 T-cells have been implicated in controlling *MTB* infection through production of interferon-γ (IFN- γ), which is critical in activating macrophages accumulating around the caseous center and engulfing and destroying escaping bacilli to prevent further tissue damage (Tufariello *et al.*, 2003). Another role for T-cells is to kill poorly activated macrophages in which the bacilli are replicating.

Disease progression occurs through liquefaction of the caseous center of necrotic lesions forming an ideal culture for the bacilli, high in oxygen (Figure 1-1(E)). The cause of liquefaction is unknown. Since macrophages are unable to survive in this condition partly due to its toxic fatty acid content, the bacilli multiple extracellularly to large numbers. Liquefied caseous material is discharged into the surrounding blood vessels and airways. Coughing aerosolizes the bacilli and large numbers are spread to the outside environment (Dheda *et al.*, 2005).

1.2 Vaccination and treatment

Immunisation with the bacilli Calmette-Guérin (BCG) is still the only vaccine available against tuberculosis. It is a live attenuated strain of *Mycobacterium bovis* and is relatively safe and inexpensive to produce. BCG is effective against childhood forms of TB

(military TB and TB meningitis) but unfortunately the precise defence provided to adults is controversial, due to its unpredictable efficacy. Controlled clinical trials have shown variability in the protection of vaccinated people from developing pulmonary tuberculosis, the most prevalent form of the disease. Possible reasons for these differences may be due to BCG related factors (strain, preparation, and administration) or host related factors (nutrition and genetics) (Moreira, 2004). In addition to this the rise in drug resistant strains of *MTB* makes the identification of new drug targets and vaccines imperative.

The majority of tuberculosis programs, to confirm diagnosis, use direct smear examination of sputum, but, if resources permit, bacteria culture testing is undertaken. Acid-fast staining is followed by examination under the microscope. One major drawback to the methods is that the diagnosis may be inaccurate and delayed due to the slow growth rate of the organism. Molecular techniques can be used to provide a more rapid, sensitive and specific procedure to test for *MTB*. The Sherlock *Mycobacteria* Identification System protocol utilises high-performance liquid chromatography; the mycolic and fatty acids extracted from the mycobacteria cell walls are species specific and yield unique chromatogram traces. Other tests employ the uses of specific nucleic acid probes, Polymerase Chain Reaction (PCR) and interferon-γ tests (Campbell and Bah-Sow, 2006; Khan and Kimerling, 2006).

There are five standard first-line drugs available to treat active *MTB*; they are isoniazid, rifampin, ethambutol, pyrazinamide and streptomycin. In order to prevent the emergence of a drug resistant *MTB* strain, at least two drugs are prescribed and the treatment must be continued for at least 6 months but with some drug regimens, the treatment lasts much longer (Khan and Kimerling, 2006).

1.3 Latency, dormancy and persistence

The terms latency and dormancy are commonly used in describing *MTB* and TB pathogenesis. In literature these words are often used interchangeably, therefore for clarity, these terms are defined below.

- Latency refers to the presence of a tuberculosis lesion which fails to produce symptoms of its presence (Gomez and McKinney, 2004).
- Dormancy is used to describe either TB disease or the metabolic state of the tubercle bacillus. TB lesions are described as active or dormant, based on whether the associated pathology is progressing or non-progressive, respectively (Gomez and McKinney, 2004).

Within a solid caseous granuloma bacilli are unable to replicate within the macrophage due to factors such as acidic pH, low oxygen concentration and the presence of toxic fatty acids but they may remain dormant there for decades (Fenton and Vermeulen, 1996). It is estimated that only 10% of affected individuals show evidence of clinical symptoms (Collins and Kaufmann, 2001). The majority develop clinical latency where symptoms are asymptomatic and non-infectious (Flynn and Chan, 2001). However, reactivation can be triggered by immunosuppression such as infection with human immunodeficiency virus (HIV). Latency is a major obstacle in trying to combat and control TB, as one-third of the world population are estimated to harbour latent *MTB* infection, which may resume growth and activate disease at any time. All conventional TB drugs are targeted towards processes involving growth and division of the bacilli and the majority are not effective against latent bacteria (Gupta and Chatterji, 2005). Hence an understanding of the mechanisms that allow these bacilli to persist in the latent state will hopefully enable the production of drugs targeted to latent bacilli.

There is limited knowledge about the stimuli that signal entry into the dormant state or that reactivate the bacilli. *In vitro* cultures of *MTB* grown under conditions of stress *i.e.* nutrient starvation (Betts *et al.*, 2002), hypoxia (Wayne and Sohaskey, 2001), low pH (Piddington *et al.*, 2000) and low concentrations of nitric oxide (Voskuil *et al.*, 2003) are thought to partly represent the *in vivo* conditions faced by dormant bacilli, persisting within granulomas.

Hypoxia is thought to be a major factor in inducing non-replicating persistence of the bacilli *i.e.* latency/dormancy. An *in vitro* model of dormancy for *MTB* has been developed where cultures of *MTB* were subjected to gradual oxygen depletion that led the bacteria into a non-replicating persistent state (Wayne and Sohaskey, 2001). Adaptations in this state, included bacteriostatis, and metabolic (switch to nitrate metabolism), chromosomal (duplication of the chromosome), and structural (a thicker multi-layer cell wall) changes to the bacilli as well as adopting resistance to antimicrobials. Supply of oxygen to bacteria in the non-replicating persistent state resulted in their resuscitation (Voskuil, 2004).

Variations of this model have been used to discover *MTB* genes that play a potential role in the development and maintenance of the non-replicating persistent state (Bacon *et al.*, 2004; Kendall *et al.*, 2004; Muttucumaru *et al.*, 2004; Park *et al.*, 2003; Voskuil *et al.*, 2004). The dormancy survival regulator protein (DosR) has been shown to induce nearly all of the *MTB* genes that respond to reduced oxygen tension (Park *et al.*, 2003). The role of DosR is discussed in more detail in section 1.6.8.

1.4 Virulence of *MTB*

The pathogen experiences numerous stresses and environmental changes during extracellular growth and after uptake by macrophages during the course of infection. Therefore, the ability of *MTB* to survive and to persist *in vivo* is a complex phenomenon involving specialised cell-surface components and multiple mechanisms that lead to adaptation and coordinated expression of bacterial virulence genes.

1.4.1 Envelope of *Mycobacteria*

The mycobacterial envelope is described as possessing three structural components: the plasma membrane, cell wall and capsule (Daffe and Etienne, 1999). Little is known about the plasma membrane of *MTB*, however, studies have implicated the cell wall and capsule in the pathogenesis of *MTB*. Both the capsule and cell wall are complex and unique, providing vital barriers against and interfaces to the host.

The *MTB* capsule contains glycans that have been shown to mediate the adhesion to and penetration of the bacilli into the hosts cells (Daffe and Etienne, 1999).The *MTB* cell wall consists of a typical lipid bilayer surrounded by a rigid peptidoglycan layer, which is covalently linked to arabinogalactan, which in turn is attached to mycolic acids. This complete structure is referred to as the mycolyl arabinogalactan-peptidoglycan and forms the core of the cell wall. Mycolic acids form a relatively impermeable layer in the envelope providing a barrier to small hydrophilic molecules and enabling resistance to chemical injury. Also associated with the cell wall are numerous free lipids, glycolipids and proteins.

Lipoarabinomannan (LAM), a major lipid of the cell wall, attached to the plasma membrane, can enhance *MTB* survival by inhibiting various antimicrobial functions of the macrophage. LAM is implicated in scavenging for toxic oxygen radicals and blocking activation of macrophages (Chan *et al.*, 1991). LAM can also bind to Toll host receptors and diffuse into their membranes causing damage to the host (Brennan, 2003). Cord Factor, another cell wall glycolipid has been shown to diffuse into host cell membranes to damage their function thereby inhibiting phagosome-lysosome fusion and promoting lung granulomas (Daffe and Etienne, 1999).

1.4.2 Gene expression regulatory mechanisms

In prokaryotes transcriptional control plays a major role in regulating gene expression in response to environmental changes. In *MTB* this is achieved by either sigma (σ) factors or two component signal transduction systems.

1.4.2.1 Sigma Factors

Sigma factors bind to core RNA polymerases to give specificity for a particular promoter. The combination of various sigma factors with RNA polymerases provides a means of modulating gene expression profiles according to changes in environmental conditions. Bacterial genomes encode a principle sigma factor, dedicated to the transcription of house keeping genes and may also contain a variable number of alternative sigma factors (Rodrigue *et al.*, 2006).

The *MTB* genome has been predicted to encode for 13 sigma factors (Cole *et al.*, 1998), named alphabetically, SigA to SigM. The functions of the sigma factors are shown in Table 1-1 and those sigma factors that are associated with virulence and adaptation are discussed below.

SigA (also named rpoV) is a principle sigma factor and is essential for bacterial viability (Gomez *et al.*, 1998; Hu and Coates, 1999). SigA was the first mycobacterial sigma factor to be associated with a virulence defect in animal models. The R515H mutant of SigA in *M.bovis* caused attenuation of virulence in the guinea pig model of infection (Collins *et al.*, 1995). The R515H mutation, located in the C-terminal portion of SigA, had no effect on driving the expression level of house keeping genes and, therefore, cell viability, but is in a region known to interact with transcriptional activators in other bacteria. A yeast two-hybrid system showed that the C-terminal portion interacts with a putative transcriptional regulator WhiB3, where the interaction is abolished in the R515H SigA mutant (Steyn *et al.*, 2002). The expression of *sigA* in some clinical TB strains has been shown to increase after phagocytosis and appears to be responsible for enhanced intracellular growth and an increased resistance to reactive oxygen intermediates (Wu *et al.*, 2004).

Located approximately 3kb downstream from the *sigA* gene is the *sigB* gene that is almost identical to the last 600bp of *sigA* (Doukhan *et al.*, 1995). SigB, like SigC, SigE, SigH and SigF plays a role in mycobacterial stress responses. A *MTB sigB* knockout mutant is more sensitive to various stress conditions including heat shock, oxidative stress and detergent (Manganelli *et al.*, 1999) but grows normally in human macrophages.

Table 1-1: Sigma factor genes in *MTB*.

Sigma Factor	Function	Reference
σA	Housekeeping and virulence	(Gomez *et al.*, 1998)
σB	Stress response	(Manganelli *et al.*, 1999)
σC	Sress response and virulence	(Thakur *et al.*, 2007)
σD	Governs the expression of a small set of ribosomal genes expressed in stationary phase during in vivo growth and virulence	(Cole *et al.*, 1998) (Calamita *et al.*, 2005)
σE	Stress response and virulence	(Wu *et al.*, 1997)
σF	Controls late-growth and virulence	(DeMaio *et al.*, 1996)
σG	Induced during the infection of human macrophages	(Cappelli *et al.*, 2006)
σH	Oxidative stress, heat shock response and virulence	(Manganelli *et al.*, 2002b)
σI	Unknown	(Cole *et al.*, 1998)
σJ	Induced in stationary phase cultures and possible role in oxidative stress	(Hu *et al.*, 2004; Hu and Coates, 2001)
σK	A mutation of *sigK* is responsible for decreased expression of MPB70 and MPB83 proteins	(Cole *et al.*, 1998) (Charlet *et al.*, 2005)
σL	Regulates polyketide synthases and secreted or membrane proteins and virulence	(Hahn *et al.*, 2005)
σM	Governs the expression of a small set of genes, including four esat-6 homologues	(Cole *et al.*, 1998) (Agarwal *et al.*, 2007)

Expression of *sigB* is dependant on several other sigma factors, including SigE (Manganelli *et al.*, 2001) and SigH (Manganelli *et al.*, 2002a). Sigma factors sigE and sigH control the expression of genes needed for the bacteria to tolerate either heat shock or oxidative stress, such as exposure to organic peroxide (Fernandes *et al.*, 1999; Manganelli *et al.*, 1999). The *M. bovis* BCG gene encoding sigF is induced after exposure to several antibiotics, oxidative stress, cold shock and starvation (DeMaio *et al.*, 1996; Michele *et al.*, 1999). However in *MTB* sigF was only induced after nutrient depletion (Manganelli *et al.*, 1999) suggesting that sigF is regulated differently in these two organisms despite their similarity.

1.5 Two component systems

Signalling systems are widespread signal transduction devices amongst prokaryotes and eukaryotes that often utilize a phosphotransfer pathway to couple an external stimulus, (*e.g.* changes in temperature, nutrients, pH, and oxygen concentrations) to a cell response (section 1.5.2). This enables rapid adaptations in reaction to changes in the environmental surroundings. In prokaryotes this normally involves a His-Asp phosphotransfer system whilst eukaryotes predominantly utilise a Ser/Thr and Tyr phosphotransfer mechanism (Stock *et al.*, 2000).

1.5.1 Domain organisation of two-component signal transduction proteins

Generic prokaryotic two-component systems involve two conserved elements. The first is the sensory protein and the second element is a response-regulator protein (RR) that controls the response, normally by altering gene expression or the activity of proteins involved. Both the sensory and RR proteins contain a wide variety of domains allowing interconnections with other regulatory systems and complex adaptive responses.

1.5.1.1 The sensory protein

The sensory protein is composed of two functional domains, an N-terminal input domain and a C-terminal transmitter domain; each of which are comprised of more than one structural domain. As the C-terminal transmitter domain contains kinase activity the sensory protein is often referred to as the histidine kinase (HK) domain.

The input domains of the sensory proteins have little sequence similarity and differ in membrane topology and domain arrangement, reflecting the numerous specific

types of environmental signals/ligands to which HKs respond, whereas the C-terminal region is highly conserved, this is depicted in Figure 1-2.

N-terminal input domain

The most common sensing modules associated with the input domain are PAS and GAF domains (Galperin *et al.,* 2001), which have a similar structural topology (Aravind and Ponting, 1997).

- PAS domains were named after three eukaryotic proteins: **P**eriod, **a**ryl-hydrocarbon receptor nuclear translocator and **s**ingle-minded, but are found in *Bacteria, Archaea* and *Eukarya*. They are important signalling modules that monitor changes in light, oxygen, redox potential and small ligands. The crystal structure of the PAS domain of the sensory protein FixL has been solved and contains a haem group. Binding of oxygen alters the coordination state of the haem iron, regulating kinase activity (Miyatake *et al.,* 2000). The three-dimensional structure of this sensing domain consists of a five-stranded anti-parallel β-barrel which appears to be conserved amongst PAS domains (Aravind and Ponting, 1997).

- GAF (cyclic **G**MP, **a**denylyl cyclase, **F**hlA) domains can bind an array of small molecules to regulate signalling events (Aravind and Ponting, 1997). Cyclic nucleotides, such as cyclic guanidine monophosphate (cGMP), are the most common ligands to bind GAF domains. Phosphodiesterases (PDEs), which bind cyclic nucleotides, contain one or two GAF domains. The crystal structure of the tandem GAF domain PDE2A reveals that each of the two GAF domains have entirely different functions. The first GAF domain binds cGMP stimulating its catalytic activity whereas the second is involved in dimerisation (Martinez *et al.,* 2002).

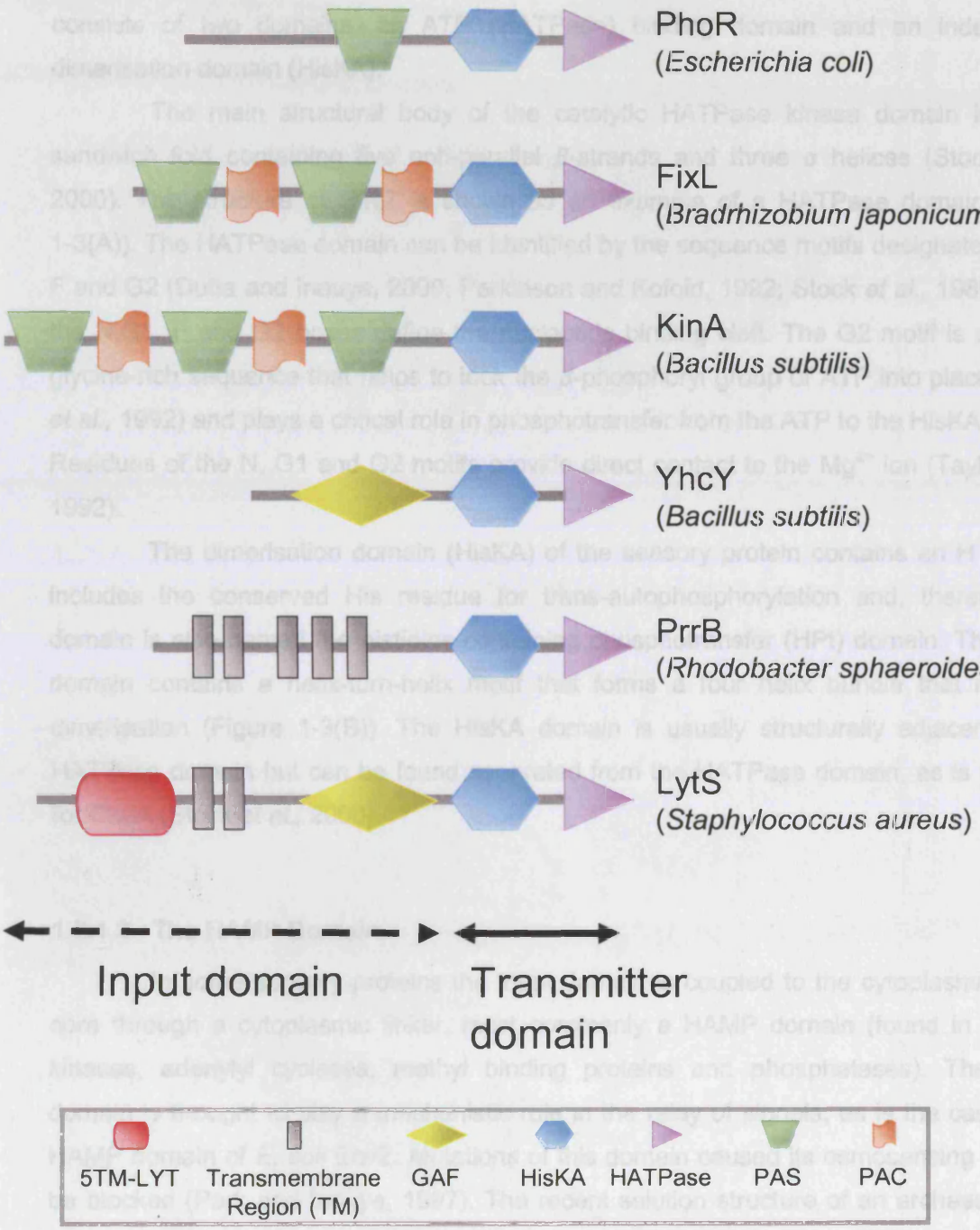


Figure 1-2: A selection of HK sensory proteins are shown to illustrate the variability of the input domain. Domains are defined using SMART (Schultz *et al.*, 1998).

C-terminal transmitter domain

The histidine kinase core within the transmitter domain is highly conserved and consists of two domains, an ATP (HATPase) binding domain and an independent dimerisation domain (HisKA).

The main structural body of the catalytic HATPase kinase domain is an $\alpha\beta$ sandwich fold containing five anti-parallel β-strands and three α helices (Stock *et al.*, 2000). The structure of EnvZ is shown as an example of a HATPase domain (Figure 1-3(A)). The HATPase domain can be identified by the sequence motifs designated N, G1, F and G2 (Dutta and Inouye, 2000; Parkinson and Kofoid, 1992; Stock *et al.*, 1989) where the N,G1, F and G2 boxes define the nucleotide binding cleft. The G2 motif is a flexible glycine-rich sequence that helps to lock the β-phosphoryl group of ATP into place (Taylor *et al.*, 1992) and plays a critical role in phosphotransfer from the ATP to the HisKA domain. Residues of the N, G1 and G2 motifs provide direct contact to the Mg^{2+} ion (Taylor *et al.*, 1992).

The dimerisation domain (HisKA) of the sensory protein contains an H box that includes the conserved His residue for trans-autophosphorylation and, therefore this domain is also named the histidine-containing phosphotransfer (HPt) domain. The HisKA domain contains a helix-turn-helix motif that forms a four helix bundle that mediates dimerisation (Figure 1-3(B)). The HisKA domain is usually structurally adjacent to the HATPase domain but can be found separated from the HATPase domain, as is the case for CheA (Stock *et al.*, 2000).

1.5.1.2 The HAMP Domain

In some sensory proteins the input domain is coupled to the cytoplasmic kinase core through a cytoplasmic linker, most commonly a HAMP domain (found in **h**istidine kinases, **a**denylyl cyclases, **m**ethyl binding proteins and **p**hosphatases). The HAMP domain is thought to play a mechanistic role in the relay of signals, as is the case of the HAMP domain of *E. coli* EnvZ. Mutations of this domain caused its osmosensing ability to be blocked (Park and Inouye, 1997). The recent solution structure of an archaeal HAMP domain shows a homodimeric, parallel four-helical coiled coil. From the structure determined, it is suggested that the HAMP domain relays a signal to the C-terminal transmitter domain via rotation of the α-helices along an axis perpendicular to the membrane (Hulko *et al.*, 2006).

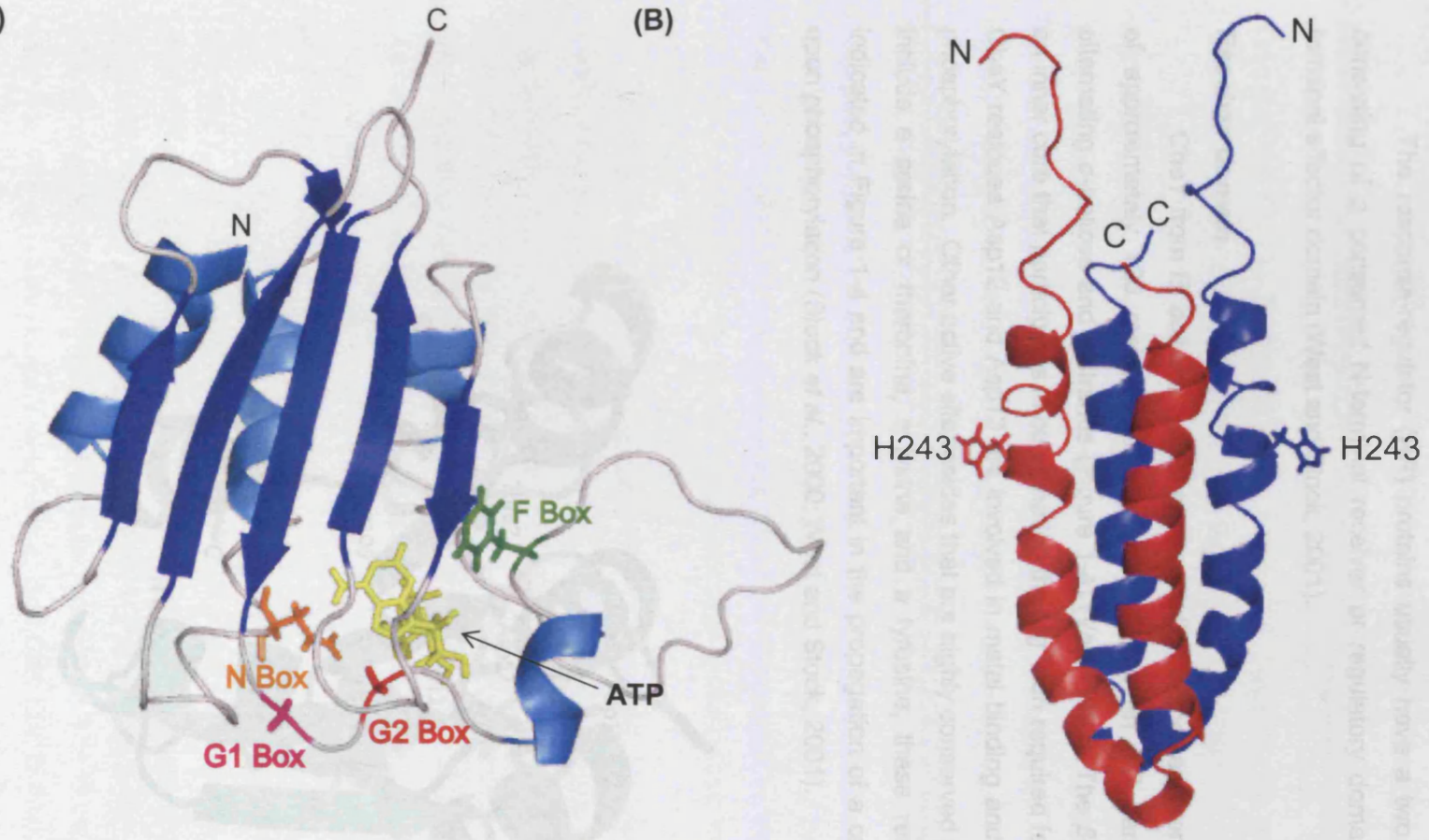


Figure 1-3: The structure of EnvZ from *E. coli* (A) HATPase domain (pdb 1BXD). Highlighted are the conserved boxes present in all catalytic domains of sensory proteins, the ATP analogue is colored yellow (B) HisKA domain (1JOY pdb), H243 is the site of autophosphorylation. Figure rendered using Pymol (DeLano, 2002).

1.5.1.3 The Response-Regulator

The response-regulator (RR) proteins usually have a two domain architecture, consisting of a conserved N-terminal receiver or regulatory domain and a variable C-terminal effector domain (West and Stock, 2001).

Receiver domain

CheY from *E. coli* is a typical representation of the receiver domain and consists of approximately 120 residues. The structure forms an α/β barrel with five sets of alternating α-helices and β-strands (Figure 1-4) (Volz, 1993). The β-strands align to form an inner core that participates in coordinating the Mg^{2+} ion required for phosphorylation. In CheY residues Asp12 and Asp13 are involved in metal binding and Asp57 is the site for phosphorylation. Other active site resides that are highly conserved in a receiver domains include a serine or theronine, a lysine and a tyrosine, these residues for CheY are indicated in Figure 1-4 and are important in the propagation of a conformational change upon phosphorylation (Stock *et al.*, 2000; West and Stock, 2001).

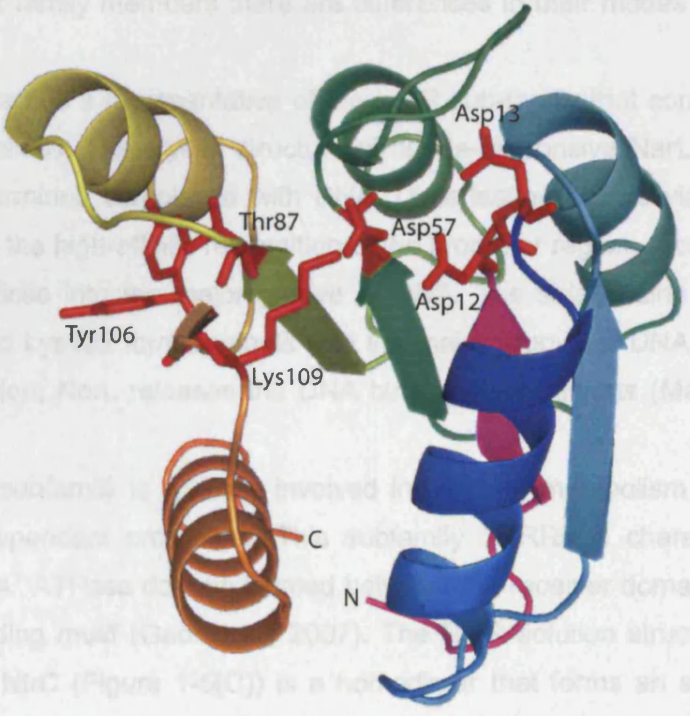


Figure 1-4: The response regulatory domain of CheY from *E. coli* (pdb 3CHY). Figure created using Pymol (DeLano, 2002).

Effector domain

RRs relay a diverse set of output responses dictated by their effector domains. The most common function of RRs is the regulation of gene expression whereby phosphorylation of the receiver domain serves to modulate the activity of DNA binding to the effector domain *i.e.* they act as transcriptional regulators. DNA binding effector domains can be sub-divided into three categories: OmpR, LuxR and NtrC-like. All three classes contain helix-turn-helix motifs but their folds are unique.

The OmpR sub-family contains a winged-helix-turn-helix binding motif of which *E. coli* OmpR is the most extensively investigated. OmpR regulates the expression of the porin genes *ompF* and *ompC* in response to changes in osmolarity, detected by the corresponding sensor EnvZ. The crystal structure of the 98 amino acid OmpR effector domain has been solved (Figure 1-5(A)). OmpR contains an N-terminal four anti-parallel β-sheet, three central α-helices and a C-terminal β hair-pin. In the central helices the α2-loop-α3 region forms the helix-turn-helix DNA binding motif with α3 acting as the recognition helix that interacts with the major grove of DNA. There are two loop 'wings' that flank the recognition helix and interact with the minor groove of DNA (Martinez-Hackert and Stock, 1997). Although there are structural similarities in the DNA binding motif between the OmpR family members there are differences in their modes of action (Stock *et al.*, 2000).

The RR NarL is a representative of the LuxR sub-family that contains a compact bundle of four α-helices. The crystal structure of nitrate-responsive NarL protein from *E. coli* has been determined complexed with DNA. Dimerisation occurs via the C-terminal helix which confers the high-affinity recognition of the promoter region, allowing insertion of the recognition helices into the major groove of DNA. The side chains of the residues Lys188, Val189 and Lys192 form contacts with the major groove of DNA (Figure 1-5(B)). Upon phosphorylation, NarL releases the DNA binding determinants (Maris *et al.*, 2005; Maris *et al.*, 2002).

The NtrC subfamily is typically involved in nitrogen metabolism and acts as an activator of σ54-dependant promoters. This subfamily of RRs is characterised by the presence of an AAA^+ ATPase domain located between the receiver domain and the helix-turn-helix DNA-binding motif (Gao *et al.*, 2007). The NMR solution structure of the DNA binding domain of NtrC (Figure 1-5(C)) is a homodimer that forms an anti-parallel four-helix fold. The structure solved was unable to bind DNA due to three point mutation in the recognition helix (Pelton *et al.*, 1999). X-ray solution scattering and EM structures of full length Ntrc RR have recently been reported (De Carlo *et al.*, 2006).

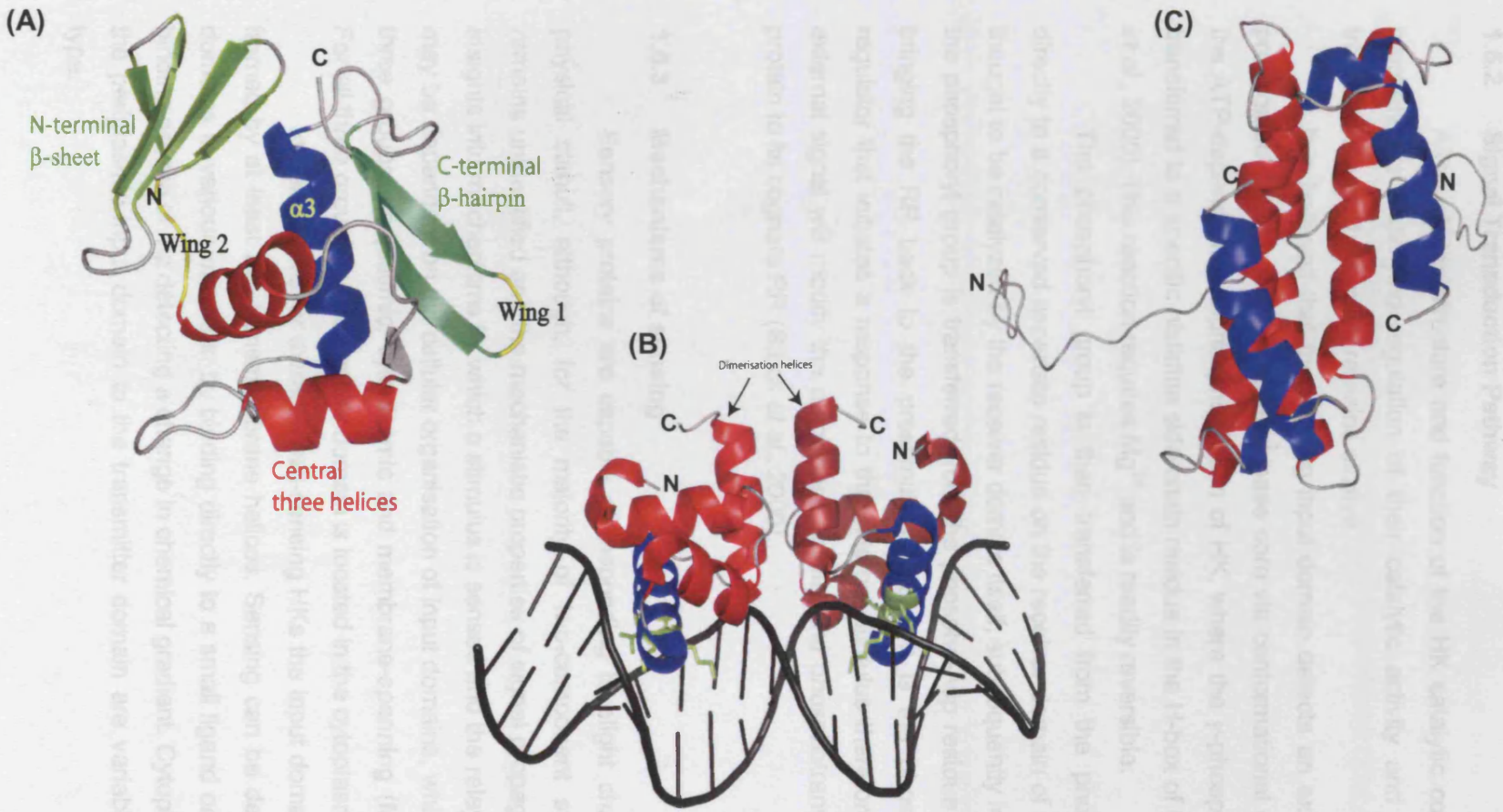


Figure 1-5: The three classes of DNA binding effector domains. The recognition helices are shown in blue. (A) *E. coli* OmpR (pdb 1OPC) (B) *E. coli* NarL complexed with DNA (pdb 1RNL), residues interacting with the major groove of DNA are shown in green (C) *Salmonella typhimurium* NtrC (pdb 1NTC). Figure generated using Pymol (DeLano, 2002).

1.5.2 Signal Transduction Pathway

Although the structure and function of the HK catalytic core is well defined the mechanisms underlying regulation of their catalytic activity and intermolecular signal transfer through the system remains elusive.

It is believed that the sensing input domain detects an external stimulus that is propagated through to the catalytic kinase core via conformational changes. This causes the ATP-dependent autophosphorylation of HK, where the γ-phosphoryl group in ATP is transferred to a specific histidine side-chain residue in the H-box of the HK domain (Stock *et al.*, 2000). This reaction requires Mg^{2+} and is readily reversible.

The phosphoryl group is then transferred from the phosphohistidine residue directly to a conserved aspartate residue on the receiver domain of a cognate RR. This is thought to be catalyzed by the receiver domain itself; subsequently in a hydrolysis reaction the phosphoryl group is transferred from the phospho-Asp residue to water (Figure 1-6) bringing the RR back to the prestimulated state. It is the phosphorylated response-regulator that initiates a response to the external stimulus therefore any changes to the external signal will modify the autokinase activity and phosphotransfer activity of the HK protein to its cognate RR (Stock *et al.*, 2000).

1.5.3 Mechanisms of sensing

Sensory proteins are capable of responding to slight changes in chemical or physical stimuli, although, for the majority of two-component systems, the stimulus remains unidentified and the mechanistic properties of signal propagation are sparse. The insights into mechanisms by which a stimulus is sensed and the relay of a signal to the RR may be dependant on the cellular organisation of input domains, which can be divided into three groups: periplasmic, cytoplasmic and membrane-spanning (Mascher *et al.*, 2006). For all three groups the transmitter domain is located in the cytoplasm.

In periplasmic (or extracellular)-sensing HKs the input domain is extracellular and framed by at least two transmembrane helices. Sensing can be detected by these input domains in various manners: by binding directly to a small ligand or a periplasmic solute-binding protein or via detecting a change in chemical gradient. Cytoplasmic linkers that join the periplasmic input domain to the transmitter domain are variable in size, extent and type.

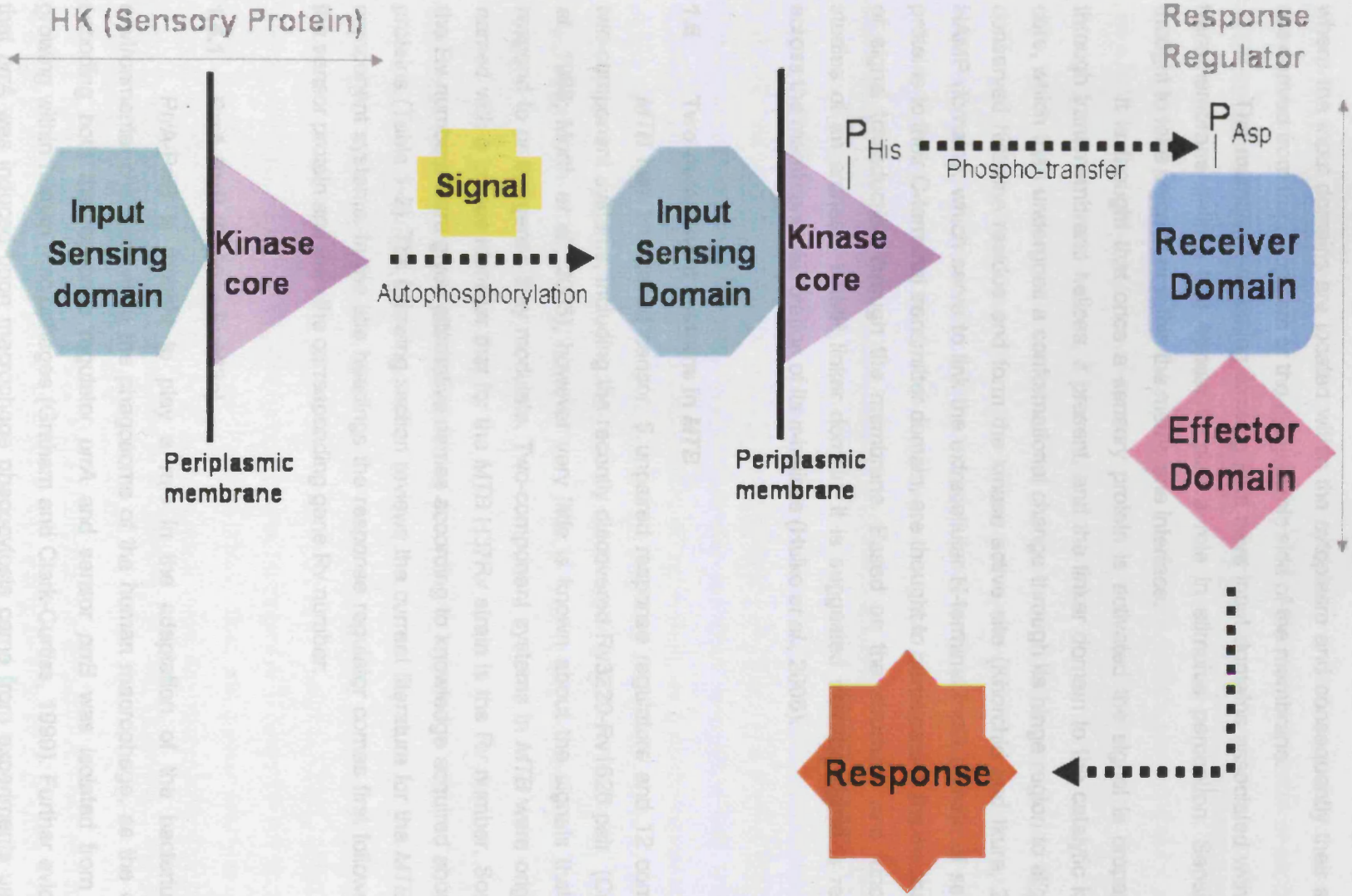


Figure 1-6: A prototypical two-component signal transduction system where P refers to the *γ*-phosphoryl group.

Cytoplasmic-sensing HKs include membrane associated or soluble proteins where the input domains are located within the cytoplasm and consequently their stimuli are derived from the cytoplasm or the cytoplasmic side of the membrane.

The membrane-spanning sensing HKs have input domains associated with 2-20 transmembrane helices that appear to play a role in stimulus perception. Sensing is thought to take place at or within the membrane interface.

It is thought that once a sensory protein is activated the signal is propagated through transmembrane helices, if present, and the linker domain to the catalytic kinase core, which then undergoes a conformational change through its hinge region to align the conserved histidine residue and form the kinase active site (Khorchid and Ikura, 2006). HAMP domains, which serve to link the extracellular N-terminal input domain of sensory proteins to their C-terminal transmitter domain, are thought to participate in the mechanism of signal transduction through the membrane. Based on the structural and functional studies of an archaeal HAMP linker domain, it is suggested that the signal is relayed across the membrane via rotation of its α-helices (Hulko *et al.*, 2006).

1.6 Two-component systems in *MTB*

MTB has 1 unpaired sensor, 5 unpaired response regulators and 12 complete two-component systems, including the recently discovered Rv3220-Rv1626 pair (Cole *et al.*, 1998; Morth *et al.*, 2005), however very little is known about the signals that they respond to or the genes they modulate. Two-component systems in *MTB* were originally named with a unique identifier that for the *MTB* H37Rv strain is the Rv number. Some of the Rv numbers were given alternative names according to knowledge acquired about the proteins (Table 1-2). The following section reviews the current literature for the *MTB* two-component systems. In the title headings the response regulator comes first followed by the sensor protein and then the corresponding gene Rv number.

1.6.1 PrrA-PrrB (Rv903c-Rv902c)

PrrA-PrrB is thought to play a role in the adaptation of the bacterium to environmental changes within the phagosome of the human macrophage, as the cDNA encoding both the response regulator *prrA* and sensor *prrB* was isolated from *MTB* growing within human macrophages (Graham and Clark-Curtiss, 1999). Further evidence that *prrA* was induced upon macrophage phagocytosis came from experiments using a transposon mutant of *prrA*, where the intracellular multiplication of *MTB* was impaired

during the first few days of mouse macrophage infection (Ewann *et al.*, 2002). Phosphorylation experiments identified autophosphoryation of His tagged PrrB protein and phosphotransfer to PrrA, establishing a sensor-receptor relationship (Ewann *et al.*, 2004). PrrA belongs to the OmpR class of DNA binding response regulators. PrrA is able to bind to its own promoter region autoregulating its expression, which was shown via DNA shift assays and observations of fluorescence with wild type and mutant *prrA-prrB MTB* strains containing a gfp reporter gene under control of the *prrA-prrB* promoter region. The precise signal(s) detected by PrrB have yet to be determined.

Recently, the crystal structure of the RR PrrA was solved complexed with either Ca^{2+} or Mg^{2+} and showed two molecules in the asymmetric unit. The N-terminal receiver domain consists of five parallel β-strands and five α-helices, as found in CheY (Figure 1-7(A)). The active site contains all the residues crucial for phosphorylation, namely Asp14, Asp15, Asp58 and Lys108. The C-terminal effector domain is composed of a four stranded anti-parallel β-sheet preceded by a three-helix bundle and a terminal β-hairpin (Figure 1-7(A)). When compared to two other known structures of the OmpR family, DrrB and DrrD, the effector domain is in a different orientation and there is a distinct difference in the transactivation loop located between α7 and the DNA recognition helix α8. The crystal structure shows the closed conformation of the protein, as a conformation change would need to occur to expose the transactivation loop and recognition helix to allow for DNA binding. Therefore, this closed conformation is postulated to be an inactive conformation of the receiver domain that is weakened upon phosphorylation (Nowak *et al.*, 2006a).

The structure of the catalytic domain (residues 338-445) of the sensory protein PrrB was solved in the absence of ATP by X-ray crystallography and conforms to known structures of HATPase kinase domains *i.e.* an $\alpha\beta$ sandwich consisting of mixed five stranded β sheet with three α helices (Figure 1-7(B)). Conserved residues were identified as being Asp353 (from the N box), Gly382, Gly384 (G1 box), Phe394 (F box), Gly407 and Gly409 (G2 Box). Residues 397-409 involved in the binding of ATP (the ATP lid) appeared to be flexible and disordered and are therefore absent from the crystal structure. In an attempt to stabilise the ATP lid, ATP analogs were used for co-crystallisation but no additional electron density was observed in the binding site. This observation was proved to be due to crystal packing (Nowak *et al.*, 2006b).

Small angle X-ray scattering of the C-terminal transmitter (HATPase and HisKA) and HAMP domains of PrrB shows that the protein exists as a dimer in solution, which is consistent with gel filtration data (Nowak *et al.*, 2006b).

Table 1-2: Two-component regulatory systems from *MTB*. Redrawn from (Rison *et al.*, 2005).

Sensor	Regulator	Number of residues	Acronym	Reference
SenX3	RegX3	410 227	SENsor X3 REGulator X3	(Wren et al., 1992)
PhoR	PhoP	485 247	By homology to Bacillus subtilis PhoR (for alkaline PHOsphatase regulation)	(Cole et al., 1998)
NarS	NarL	425 216	By homology to NarL proteins; NitrAta Reductase; narS coined by Parish *el al.* to distinguish narS from a probable nitrate reductase gene, also known as *narX*	(Parish et al., 2003a)
PrrB	PrrA	446 236	Phagocytosis Response Regulator	(Graham and Clark-Curtiss, 1999)
MprB	MprA	504 230	Mycobacterial Persistence Regulator	(Zahrt and Deretic, 2001)
KpdD	KdpE	860 226	By homology to E. coli kdpED (from Potassium DePendence, (Epstein and Davies, 1970)	(Cole et al., 1998)
TrcS	TrcR	509 257	Tuberculosis Regulatory Component	(Haydel et al., 1999a)
TcrY	TcrX	475 234	Two-Component Regulator	(Parish et al., 2003a)
DosS (DevS)	DosR (DevR)	578 217	dev: Differentially Expressed in the Virulent strain (H37Rv vs. H37Ra) dos: DOrmancy Survival or Dependent on Oxygen Status	(Boon and Dick, 2002; Kinger and Tyagi, 1993; Rison et al., 2005)
MtrB	MtrA	567 228	*Mycobacterium tuberculosis* Response regulator	(Via et al., 1996)
TcrB2	TcrA	168 253	Two-Component Regulator	(Shrivastava et al., 2006)
Rv3220	Rv1626	501 205	N/A – function unknown	(Cole et al., 1998)
dosT		573	cf. *dosRS*	(Roberts et al., 2004)
	Rv0195	211	N/A – function unknown	(Haydel and Clark-Curtiss, 2004)
	Rv0260c	381	N/A – function unknown	(Cole et al., 1998)
	Rv0818	255	N/A – function unknown	(Cole et al., 1998)
	Rv2884	252	N/A – function unknown	(Cole et al., 1998)
	Rv3143	133	N/A – function unknown	(Cole et al., 1998)

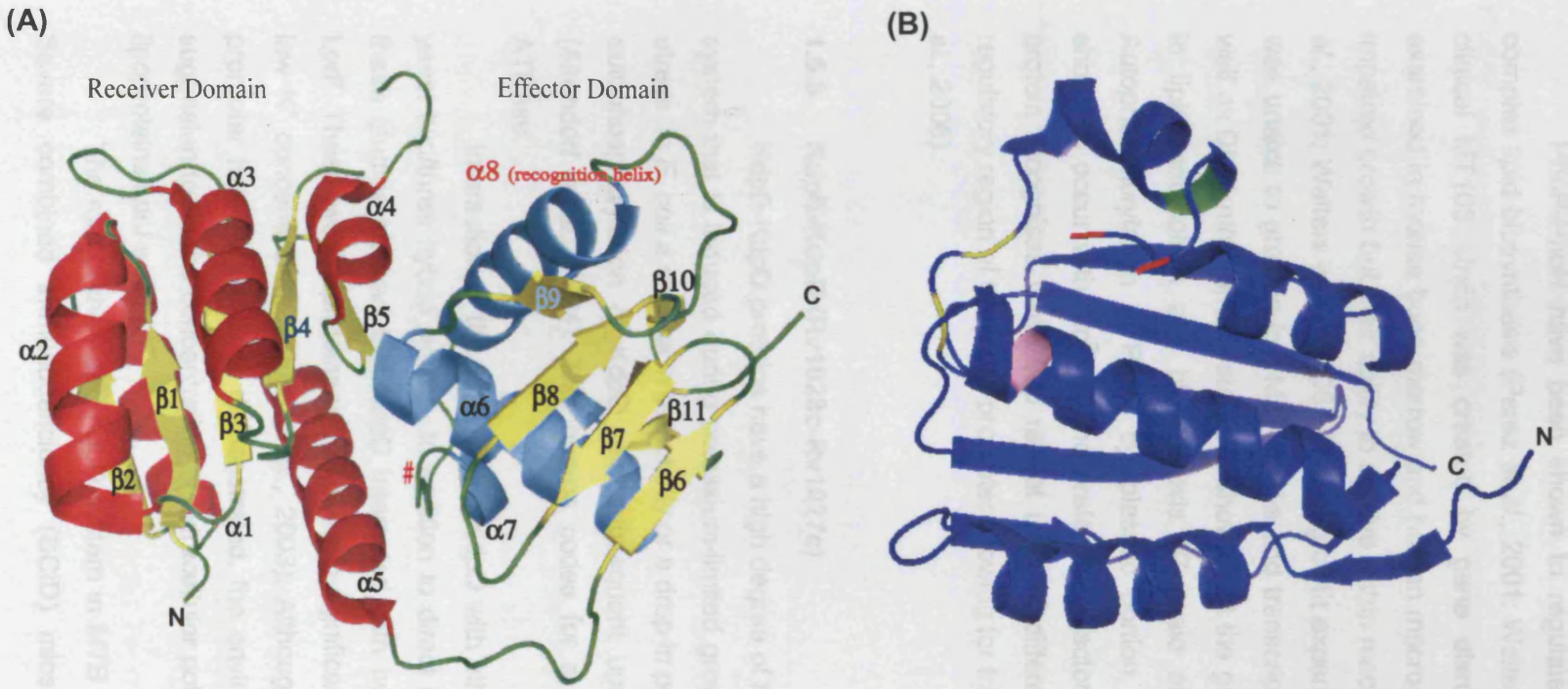


Figure 1-7 The crystal structures of the *MTB* 2CS PrrA-PrrB (A) RR PrrA (pdb 1YS6). The labeling of the α-helices, β-strands and location of the transactivation loop (#) are indicated and (B) The HATPase/catalytic domain of PrrB (pdb 1YSR). Highlighted are locations of the conserved residues F394 (green), Asp353 (pink), Gly382, and Gly384 (yellow). Location of the residues missing (397-409) from the crystal structure are shown in red. Figures created using Pymol (DeLano, 2002).

1.6.2 PhoP-PhoR (Rv0757-Rv0758):

PhoP-PhoR have been shown to regulate genes essential for virulence and complex lipid biosynthesis (Perez *et al.,* 2001; Walters *et al.,* 2006). A *phoP* mutant *MTB* clinical MT103 strain was created by gene disruption and its characteristics were examined in mouse bone marrow and human macrophages *in vivo.* The mutant exhibited impaired growth but their ability to survive within macrophages was not affected. (Perez *et al.,* 2001; Walters *et al.,* 2006). *In vitro* growth experiments demonstrated that the mutant was unable to grow in low Mg^{2+} media and transcription profiles generated from RNA as well as DNA micro-array data have shown that the genes regulated by PhoP are involved in lipid metabolism and biosynthesis (Gonzalo *et al.,* 2006; Walters *et al.,* 2006). Autophosphorylation of PhoR cytoplasmic portion and phosphotransfer to PhoP was shown to occur with Mn^{2+} as the preferred cofactor. PhoP, a site-specific DNA binding protein recognises three 9bp repeat units of different nucleotide sequences within the regulatory region of the *phoP* promoter allowing for transcriptional autoregulation (Gupta *et al.,* 2006).

1.6.3 KdpE-KdpD (Rv1028c-Rv1027c)

KdpE-KdpD proteins have a high degree of homology with the *E. coli* KdpE-KdpD system that is induced during potassium-limited growth and is also affected by oxidative stress. In *E. coli* a change in osmolarity or a drop in potassium concentration will cause the autophosphorylation of KdpD and subsequent upregulation of the *kpdFABC* operon (Altendorf *et al.,* 1994). This operon codes for a four subunit K^+ transporting P-type ATPase.

Interaction of the *MTB* sensor KdpD with other proteins was investigated using a yeast two/three hybrid system. In addition to direct interaction between KdpD and KdpE these studies showed that KdpD interacted with two membrane lipoproteins, LprJ and LprF. These two lipoproteins were shown to significantly enhance *kdpFABC* expression at low K^+ concentrations (Steyn *et al.,* 2003). Although binding of KdpE to the *kdpFABC* promoter has not yet been demonstrated, the environmental signal sensed by KdpD is suggested to be the concentration of extracellular potassium with a response mediated by lipoproteins LprJ and LprF.

The role of the KdpE-KdpD system in *MTB* pathogenesis has been investigated. Severe combined immunodeficiency (SCID) mice infected with a mutant knockout

succumbed to infection more rapidly than those infected with wild-type *MTB* (Parish *et al.*, 2003a).

1.6.4 MprA-MprB (Rv981-Rv982)

MprA-MprB is grouped into the *E. coli* EnvZ-OmpR subfamily of two-component signalling systems and has been shown to function as an intact signal-transducing pair *in vitro* and *in vivo*. The key phosphorylation residues have been identified by site-directed mutagenesis as His249 in MprB and Asp48 in MprA. Mutations introduced at these sites caused attenuation in the ability of *M. bovis* BCG to grow intracellularly in macrophages (Zahrt *et al.*, 2003); the same group demonstrated that this system is required for entrance into and maintenance of persistent infection (Zahrt and Deretic, 2001).

MprA-MprB has been shown to respond to several environmental stress conditions (including changes in pH, alkaline and nonionic detergents) and regulates two key stress responsive sigma factors, σB and σE (He *et al.*, 2006).

1.6.5 RegX3-SenX3 (Rv0491-Rv0490)

SenX3 can catalyse autophosphorylation in the presence of Mg^{2+} ions and undergo a phosphotransfer reaction with its regulator RegX3 *in vitro*. Site-directed mutagenesis identified His167 and Asp62 as the sites of phosphorylation in SenX3 and RegX3, respectively (Himpens *et al.*, 2000).

The SenX3-RegX3 system is required for virulence, which was shown when the *regX3-senX3* system was deleted from an *MTB* strain via homologous recombination. The mutant strain had subtle growth defects after infection of macrophages and was attenuated in both activated and resting macrophages and in immunodeficient and immunocompetent mice (Parish *et al.*, 2003b). Reduced virulence was also shown when knockouts of *senX3* or *regX3* were constructed and infected into mice (Rickman *et al.*, 2004).

The sensor SenX3 has been predicted to contain a PAS-like domain, which in other sensors is involved in oxygen and redox sensing. Therefore, it is postulated that this system is required for the protective response to oxidative stress in the initial stages of infection when phagocytosed by macrophages (Rickman *et al.*, 2004).

1.6.6 TrcR-TrcS (Rv1033c-Rv1032c)

Overexpression of the full-length TrcS protein in *E. coli* resulted in cell lysis, which was attributed to a transmembrane region as lysis only occurred when this region was present. The cytoplasmic portion of the sensor has been demonstrated to undergo autophosphorylation and directly phosphorylates its response regulator TrcR *in vitro* (Haydel *et al.*, 1999).

TrcR is homologous to the OmpR class of DNA binding proteins and has been identified to bind to its own promoter and induce its own expression via interaction with AT rich regions of DNA. Autoregulation was shown by co-transforming a *trcR-lacZ* fusion plasmid and a TrcR expression vector into *E. coli*. Once the TrcR protein was induced the activity of β-galactosidase from the *trcR-lacZ* fusion increased >500 fold. DNA gel shift assays confirmed that the TrcR was autoregulating via directly binding to an AT rich DNA piece of its own *trcR* promoter (Haydel *et al.*, 2002). Protein-DNA binding experiments also demonstrated that TrcR binds to a 69bp AT rich sequence within the protein Rv1057 and transcriptional analysis revealed that TrcR represses its transcription. Rv1057 is thought to consist of a repeating domain that forms a β-propeller structure; other proteins with this fold have been implicated in the pathogenesis of several human diseases. It was postulated that Rv1057 has a role in intraphagosomal survival and metabolism of lipid substrates, as it is induced upon *MTB* growth in the presence of free fatty acids as the only carbon source (Haydel and Clark-Curtiss, 2006).

Selective capture of transcribed sequences (SCOTS) experiments with cDNA probes have demonstrated that the *trcR* and *trcS* genes are expressed in early exponential phase and therefore it is postulated the genes are important in the adaptation to the intracellular environment within macrophages and in the transition from latency to reactivation (Haydel *et al.*, 2002).

Parish and co-workers have shown that an 882bp deletion in *MTB trcS* causes hypervirulence in SCID mice (Parish *et al.*, 2003a), whereas a transposon mutant of *trcS* did not show any phenotype in mouse bone marrow-derived macrophages (Ewann *et al.*, 2002).

1.6.7 MtrA-MtrB (Rv 3246c-Rv3245c)

MtrA has been shown to participate in a phosphotransfer reaction with a heterologous histidine kinase sensor CheA *in vitro* (Via *et al.*, 1996) and is essential as mutants lacking the *mtrB* gene or the *mtrA-mtrB* system are unable to be isolated (Parish *et al.*, 2003a; Zahrt and Deretic, 2000). Through the monitoring of *mtrA-gfp* fusion activity,

it was demonstrated that the *mtrA* gene is differentially expressed in virulent and avirulent strains during growth in macrophages (Zahrt and Deretic, 2000).

Increased expression of phosphorylation competent MtrA results in attenuated growth in macrophages, mice, lungs and spleens, whereas elevated levels of phosphorylation defective MtrA lead to proliferation in macrophages but attenuated growth in mice, lungs and spleens.

1.6.8 DosR-DosS (Rv3133c-Rv3132c)

The DosR-DosS system is the best characterised two-component system from *MTB.* It was originally named DevR-DevS as the two co-transcribed genes had higher expression in the virulent H37Rv *MTB* strain compared to its avirulent derivative H37Ra (Dasgupta *et al.,* 2000). This system was renamed to dormancy survival system due to its association with hypoxia and latency.

Whole genome microarray data identified genes whose expression profile was rapidly altered upon exposure to hypoxic conditions and nitric oxide. Among these genes was an operon that contained the DosR-DosS two-component system. An up-stream gene was then disrupted in order to prevent expression of the operon, which caused the hypoxic regulatory expression of α-crystallin HspX (acr) to be abolished; acr is a major cell wall protein in stationary phase mycobacteria and has been associated with dormancy (Sherman *et al.,* 2001). Protein-DNA binding experiments identified DosR to have two binding sites in the *hspX* promoter region and mutational analysis indicated that both were important for the induction of *hspX* expression, with the downstream site being essential (Park *et al.,* 2003). However deletion of the DosS gene by gene replacement from H37Rv had no effect on *acr* induction, possibly due to the presence of a homologous sensory protein DosT (Rv2027c) that may complement the function of DosS (Sherman *et al.,* 2001). Further gene expression data comparing wild-type and DosR mutant *MTB* strains indicate that DosR is a transcription factor needed to induce most of the genes that respond to hypoxia or nitric oxide (Park *et al.,* 2003; Roberts *et al.,* 2004; Sherman *et al.,* 2001; Saini *et al.,* 2004a; Voskuil *et al.,* 2003).

DosR *MTB* induces a regulon of nearly 50 genes (Sherman *et al.,* 2001). However, more than two-thirds of these induced genes have unknown function, highlighting that the *MTB* adaptation to hypoxia is not well described. Other genes, with known function includes *narX* a fused nitrate reductase, *nark2* a nitrile extrusion protein and *fdxA* encoding ferredoxin, but their roles in hypoxia have not yet been determined.

The DosR-DosS two-component system is also expressed following treatment with ethanol or being moved from a rolling to a standing culture (Kendall *et al.*, 2004).

In vivo studies that sought to mimic caseous granulomas, by implanting artificial granulomas containing encapsulated *MTB* within hollow fibers, demonstrated through gene and mutant expression survival patterns that DosR mediates the transition into dormancy (Karakousis *et al.*, 2004). Therefore, the DosR-DosS system has been postulated to be crucial in adaptation of the bacilli to hypoxic and low NO conditions experienced during latency.

Contradiction can be found in mutational studies. Through disruption of the *dosR* gene it has been suggested that DosR is required for virulence in guinea pigs but is not essential for the entry, survival or replication of the bacillus within human monocytes *in vitro* (Malhotra *et al.*, 2004), whereas a *dosR* deletion mutant showed hypervirulence in SCID mice. Parish and co-workers also suggest that *dosR* induces genes that may slow down the replication of bacilli *in vivo* and/or increase its resistance to host killing mechanisms (Parish *et al.*, 2003a).

Full length DosS (DevS) is predicted to encode a 578 amino-acid protein that contains two GAF domains (residues 63-205 and 231-377) and three transmembrane segments (between residues 187-203, 274-295 and 325-342) within its input sensing domain. The cytoplasmic portion (residues 358-378) containing the kinase core was cloned with a His tag and overexpressed in *E. coli* to produce insoluble protein that was refolded on Ni^{2+}-NTA resin. Refolded DosS was shown to participate in autophosphorylation at His395 and rapid phosphotransfer to Asp54 of DosR in an Mg^{2+} dependant manner. Mutations in Asp8, Asp9 and Lys104 of DosR and Asn503 of DosS abolished phosphoryl transfer (Roberts *et al.*, 2004; Saini *et al.*, 2002; Saini *et al.*, 2004a). These experiments established DosS and DosR as being a typical cognate sensor-regulator pair.

DosR belongs to the NarL family of response regulators; the C-terminal effector domain of DosR (residues 144-217) has been solved by crystallography and consists of four α-helices that assemble into tetramers forming a dimer in the asymmetric unit (Figure 1-8(A)). It was also co-crystallised with a 43mer DNA duplex containing the 20bp consensus promoter of hypoxia-induced genes (Figure 1-8(B)). Where the side chains of the residues Lys179, Lys182 and Asn183 form contacts with the major groove of DNA (Wisedchaisri *et al.*, 2005). The DNA sequence contains inverted repeats of a G4G5G6A7C8T9 recognition motif where G4G5G6 has a distorted β-DNA structure and is thought to be the place for promoter recognition by DosR.

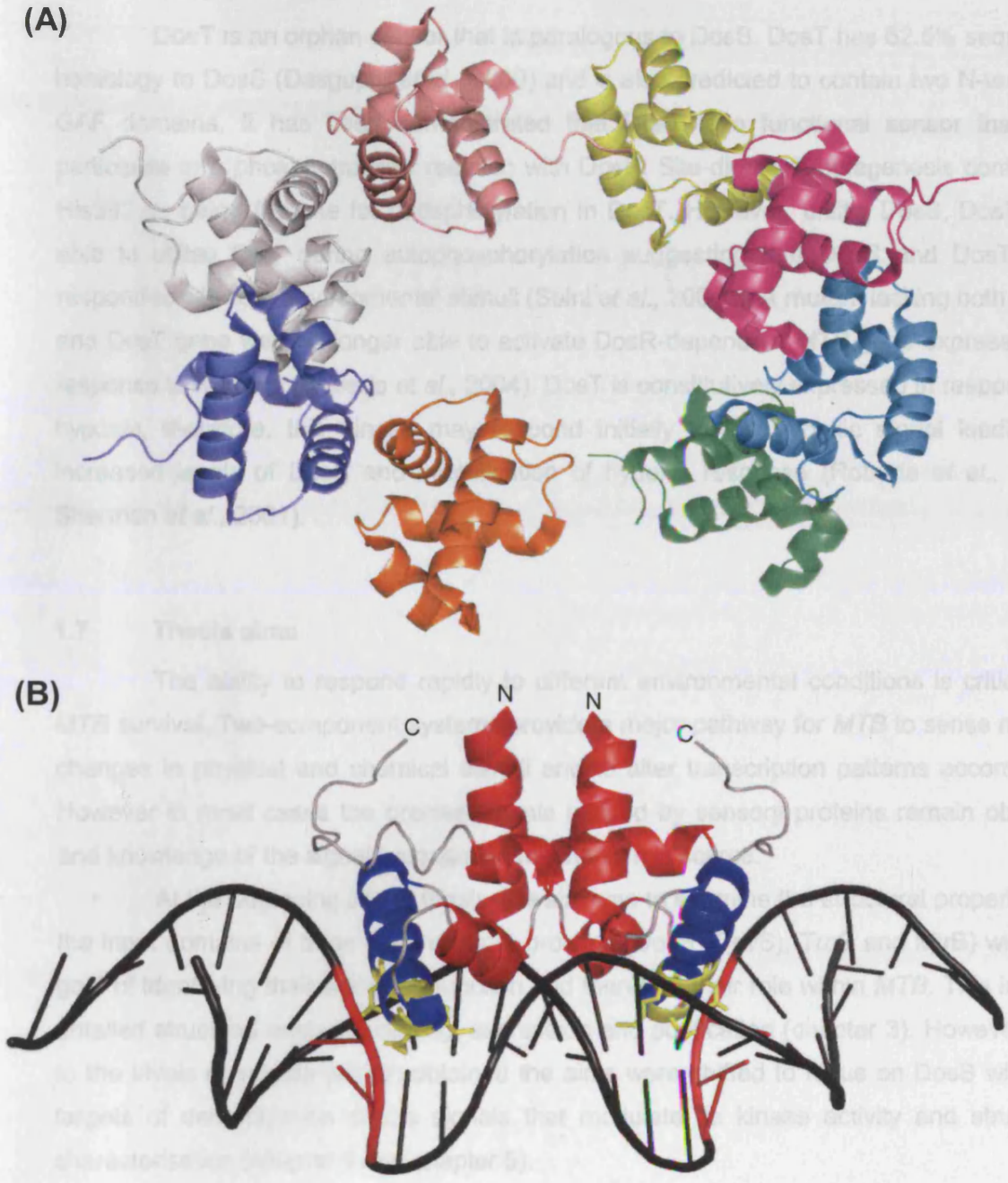


Figure 1-8: Crystal structures of *MTB* DosR (A) the effector domain forming a dimer each consisting of four tetramers (pdb 1ZLJ) (B) the effector domain complexed with DNA (pdb 1KLK). The recognition helices are shown in blue, the residues interacting with the major groove of DNA are colored yellow and the recognition motif of the DNA is indicated in red.

1.6.9 DosT (Rv2027c)

DosT is an orphan sensor that is paralogous to DosS. DosT has 62.5% sequence homology to DosS (Dasgupta *et al.*, 2000) and is also predicted to contain two N-terminal GAF domains. It has been demonstrated that DosT is a functional sensor that can participate in a phosphotransfer reaction with DosR. Site-directed mutagenesis confirmed His392 as being the site for phosphorylation in DosT. However, unlike DosS, DosT was able to utilise Ca^{2+} during autophosphorylation suggesting that DosS and DosT may respond to different environmental stimuli (Saini *et al.*, 2004b). A mutant lacking both DosS and DosT gene was no longer able to activate DosR-dependant *MTB* gene expression in response to hypoxia (Roberts *et al.*, 2004). DosT is constitutively expressed in response to hypoxia, therefore, this kinase may respond initially to the hypoxic signal leading to increased levels of DosS and amplification of hypoxic response (Roberts *et al.*, 2004; Sherman *et al.*, 2001).

1.7 Thesis aims

The ability to respond rapidly to different environmental conditions is critical for *MTB* survival. Two-component systems provide a major pathway for *MTB* to sense minute changes in physical and chemical stimuli and to alter transcription patterns accordingly. However in most cases the precise signals sensed by sensory proteins remain obscure and knowledge of the signal propagation mechanism is scarce.

At the beginning of this thesis, the aim was to examine the structural properties of the input domains of three *MTB* sensory proteins (DosS (DevS), TrcS and MtrB) with the goal of identifying their signalling function and therefore their role within *MTB*. This initially entailed structural analysis, cloning, expression and purification (chapter 3). However due to the levels of soluble protein obtained the aims were shifted to focus on DosS with the targets of determination of the signals that modulate its kinase activity and structural characterisation (chapter 4 and chapter 5).

CHAPTER TWO

MATERIALS AND METHODS

2 Materials and Methods

Chapter two outlines the experimental protocols and the techniques routinely employed in this work. All reagents were purchased from Sigma-Aldrich or VWR unless otherwise stated.

2.1 Gene annotation

In order to identify domains, transmembrane helices and signal peptides the full length sequence of four *MTB* sensory proteins (DosS, DosT, TrcS and MtrB) were submitted to the following databases: **TM-HMM** (transmembrane helices based on a hidden Markov model, http://www.cbs.dtu.dk/services/TMHMM) (Krogh *et al.*, 2001), **SMART** (simple modular architecture research tool, http://smart.embl-heidelberg.de/) (Schultz *et al.*, 1998), **Pfam** (protein families database of alignments and hidden Markov models, http://www.sanger.ac.uk/Software/Pfam/) (Bateman *et al.*, 2004) and **SignalP** (signal peptide, http://www.cbs.dtu.dk/services/SignalP) (Nielsen *et al.*, 1997). The full length DosS sequence was also submitted to a motif search database, PPsearch (http://www.ebi.ac.uk/ppsearch/, European Bioinformatics Institute).

2.2 Bacterial strains

The bacterial strains used in this study are either *Escherichia coli* DH5α or *Escherichia coli* XL1-blue for cloning of microbacterial DNA and either *Escherichia coli* BL21 (DE3) pLysS, *Escherichia coli* Rosetta (DE3) pLysS, *Escherichia coli* C41 (DE3) or *Escherichia coli* C43 (DE3) for the expression of proteins.

2.2.1 Competent *E. coli* cells

A single colony of BL21 (DE3) pLysS, or Rosetta (DE3) pLysS or C41 (DE3) or C43 (DE3) cells were inoculated into 5 ml Luria-Bertani (LB) media (section 2.3.1) and incubated overnight at 37 °C and 200 revolutions per minute (rpm). For the preparation of competent Rosetta (DE3) pLysS cells the LB media (section 2.3.1) contained 33 μg/ml chloramphenicol. 1 ml of the overnight culture was then used to inoculate 100 ml LB and incubated at 37 °C and 200 rpm until mid log phase (OD_{600nm} = 0.6). The cells were pelleted by centrifugation at 2000 rpm for 10 minutes at 4 °C. The supernatant was

discarded and the cell pellet resuspended in 50 ml of ice cold 100 mM $CaCl_2$ and held on ice for 1-2 hours. The cells were pelleted as previously described and resuspended in 4 ml of ice cold 100 mM $CaCl_2$ containing 15 % (v/v) sterile glycerol. Cells were aliquoted into 50 µl samples and stored at -80 °C.

E. coli DH5α or *E. coli* XL1-blue cells for cloning of microbacterial DNA were commercially bought from Novagen.

2.3 Growth medium and culture conditions

All media and media components were produced using deionised water from an Elga Maxima Ultra-pure water purification system and sterilized by autoclaving. All growth media for *E. coli* DH5α and BL21 (DE3) strains harbouring any pET30aTEV expression vector (Novagen) was supplemented with kanamycin at 50 µg/ml concentration. All growth media for *E. coli* Rosetta (DE3) pLysS strain containing any pET30aTEV vector was supplemented with kanamycin and chloramphenicol at 50 µg/ml and 33 µg/ml concentrations respectively. *E. coli* strains cultures harbouring a pAC15bTEV vector were supplemented with chloramphenicol at 33 µg/ml concentration. All liquid cultures were grown in a shaking incubator at 37 °C (unless otherwise stated).

2.3.1 Luria-Bertani (LB) growth media

For liquid cultivation of *E. coli* strains LB broth was used (consisting of 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl). LB agar plates consisted of LB medium supplemented with 15 g/L agar poured evenly into Petri dishes.

2.3.2 Minimal media

The expression of isotopically enriched proteins was carried out in M9 media, consisting of 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4 and 0.5 g/L NaCl. For the expression of [^{15}N]-labelled proteins, uniformly [^{15}N]-labelled $(NH_4)_2SO_4$ (Cambridge Isotopes Laboratories) was used at 1 g/L; for [^{13}C]-labelled proteins, uniformly labelled glucose (Cambridge Isotopes Laboratories) was used at 2 g/L. This solution was pH adjusted to 7.4 and autoclaved. The following sterile filtered solutions were added per litre of media just before inoculation: 1 ml 1 M $MgCl_2$, 5 µl 1 M $CaCl_2$, 500 µl 0.01 M $FeSO_4$, 500 µl 1000X vitamins solution*, 500 µl x1000 micronutrient solution** and 500 µl 50 mg mL^{-1} kanamycin.

Per Litre:

*1000X vitamins solution
0.4 g Choline chloride
0.5 g Folic acid
0.5 g Pantothenic acid
0.5 g Nicotinamide
1.0 g Myo-inositol
0.5 g Pyridoxal HCl
0.5 g Thiamine HCl
0.05 g Riboflavin
1.0 g Biotin

**1000X micronutrient solution
3 μM Ammonium Molybdate
400 μM H_3BO_3
30 μM $CoCl_2$
10 μM $CuSO_4$
80 μM $MnCl_2$
10 μM $ZnSO_4$

2.4 Nucleic acid manipulation

2.4.1 Purification of plasmid DNA

Plasmid DNA was prepared from freshly transformed *E. coli* DH5α cultures grown overnight in 5 ml LB media containing antibiotic at 37 °C and 200 rpm. Purification was performed using a QIAprep Spin Miniprep kit (Qiagen) following the manufactures protocol. The bacterial cells are lysed under alkaline conditions, and the lysate is subsequently neutralised and adjusted to high-salt binding conditions. The spin columns contain a unique silica membrane that ensure only DNA is adsorbed, while RNA, cellular proteins and metabolites are not retained by the membrane but are found in the flow-through. Salts are removed by a wash step and the purified plasmid eluted from the QIAprep spin column using Elution Buffer (EB, Tris containing buffer, pH 7.0 – 8.5).

2.4.2 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate and identify digested DNA fragments. DNA fragments <1 Kilobase (kb) or >1 kb in size were resolved on 1.5 % or 1 % (w/v) agarose gel, respectively. Agarose was dissolved in TAE buffer (40 mM 2-amino-2-hydroxymethyl-1,3-propanediol (Tris)-Acetate and 1 mM ethylene-diamine-tetraacetic acid (EDTA)) by heating and, after cooling, ethidium bromide was added to a final concentration of 0.5 µg/ml to visualize the DNA. Gels were cast and set in the apparatus provided. DNA samples were prepared by the addition of DNA loading buffer (0.1 % bromophenol blue, 20 mM Tris and 20 % glycerol). The 100 bp and 1 kb ladders (New

England Biolabs, NEB) were used as molecular weight standards. Electrophoresis was carried out in TAE buffer at a constant voltage of 100 V and the resolved DNA bands visualized on UV-transilluminators.

2.4.3 Extracting and purifying DNA from agarose gels

DNA fragments of the appropriate size were extracted from the gel using a clean scalpel and purified using the QIAquick gel extraction kit (Qiagen) according to the manufactures protocol. Provided, with each kit, are buffers that optimise the efficient recovery of DNA and removal of contaminants. The binding buffers provide the correct salt concentration and pH for adsorption of DNA to the silica-membrane of the spin column. Salts are quantitatively washed away by the ethanol-containing PE buffer. The DNA is subsequently eluted with buffer EB (10mM Tris, pH 8.5).

2.4.4 Constructs and expression vector

Oligonuclotide primers were designed in order to amplify various gene fragments by polymerase chain reaction (PCR). Constructs produced for each protein are shown in Table 2-1, along with their primers. The primers were engineered to be at least 16 base pairs long with similar melting points for the forward and reverse primers, and included a stop codon in the reverse primer. Primers for $DosS_{COM}$ and $DosT_{COM}$ were designed to give the gene 1 kb flanking regions on either side and encompass a BamHI (GGATCC) site at the 5' end and a HindIII (AAGCTT) site at the 3' end. The primers for constructs of DosS (except $DosS_{COM}$), TrcS and MtrB were designed to introduce cleavage sites, NcoI (CCATGG) and HindIII (AAGCTT) at the 5' and 3' ends of the amplified fragment, respectively. The amplified NcoI-HindIII and BamHI-HindIII gene products were cloned into a pET30aTEV expression vector (Novagen) with an N-terminal polyhistidine affinity tag and a tobacco etch virus (TEV) protease cleavage site to allow the removal of the affinity tag.

For DosT constructs (except $DosT_{COM}$) the primers introduced NdeI (CATATG) and BamHI (GGATCC) restriction sites at the 5' and 3' ends of the amplified fragment, respectively, and the amplified NdeI-BamHI gene products were cloned into a pAC15bTEV expression vector (a gift from the Bloomsbury Centre for Structural Biology), with an N-terminal polyhistidine affinity tag and a TEV protease cleavage site. All primers were purchased from MWG.

2.4.5 Amplification of genes using Polymerase Chain Reaction (PCR)

PCR amplification was performed in 50 µl reaction volumes containing 1x *Pfu* DNA polymerase buffer (Promega) (200 mM Tris HCl pH 8.8, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1.0 % TritonX-100 and 1 mg mL^{-1} nuclease-free BSA), 200 µM of each deoxynucleoside triphosphates, 1 µM of upstream primer, 1 µM of downstream primer, 0.5 µg DNA template and 2.5 units (U) *Pfu* polymerase enzyme (Promega). As a DNA template, a cosmid library of *Mycobacterium Tuberculosis* (*MTB*) was used. This library was created in the laboratory of our collaborator Professor Neil Stoker at the Royal Veterinary College. PCR reactions were carried out using the Primus thermocycler (MWG) set to hold for 95 °C for 5 minutes before performing 30 cycles of 95 °C for 1 minute, annealing temperature for 1 minute and a 72 °C extension for 1 minute/ Kbp to amplify the required stretch of DNA. Annealing temperatures are shown in Table 2-1. The PCR amplified products were analysed on an agarose gel, and purified using QIAquick gel extraction kit (Qiagen).

2.4.6 Restriction enzyme digestion

All restriction enzymes were obtained from NEB. The pET30aTEV plasmid was double digested with either NcoI or BamHI and HindIII. All DosS (except $DosS_{COM}$), MtrB, and TrcS PCR fragments were double digested with NcoI and HindIII. The pAC15b plasmid and DosT (except $DosT_{COM}$) PCR fragments were double digested with NdeI and BamHI. $DosS_{COM}$ and $DosT_{COM}$ PCR fragments were double digested with BamHI and HindIII. All digests were performed using the manufactures protocol. Reactions were performed in 20-50 µl volumes and incubated at 37°C for 90 minutes. Digested products were then electrophoresed on an agarose gel and subsequently extracted and purified using QIAquick gel extraction kit (Qiagen; section 2.4.3).

Table 2-1: Primers and Annealing temperatures used in the cloning experiments are listed below; the introduced restriction sites are shown in capitals.

Constructs	Sequences of Primers: 5'- 3'		Annealing Temperatures (^{O}C)
	Forward	Reverse	
$DosS_{1-579}$	cgg tCC ATG Gcc atg aca	a gtg AAG CTT aga gct act gcg aca a	60
$DosS_{1-379}$	cgg tCC ATG Gcc atg aca	cta AAG CTT tca gcg cat ccg acg	55
$DosS_{63-379}$	gta gCC ATG Gac ctg gag gca acc	cta AAG CTT tca gcg cat ccg acg	60
$DosS_{63-210}$	gta gCC ATG Gac ctg gag gca acc	cca cga AAG CTT cta ctt agc ctg	55
$DosS_{231-379}$	gta gCC ATG Gaa ccc gcg acg gtg	cta AAG CTT tca gcg cat ccg acg	60
$DosS_{231-470}$	gtt gtt gtc gCC ATG Gaa ccc gcg acg gtg	gct ggt gcg aag gtt tca gtc ggc aaa ttg gga	51
$DosS_{243-379}$	gtc gcc gCC ATG Gcg ctc aaa ctg acg	cta AAG CTT tca gcg cat ccg acg	68
$DosS_{250-379}$	cg gCC ATG gac gct gcc ctg gta gcc	cta AAG CTT tca gcg cat ccg acg	68
$DosS_{12-379}$	gaa aCC ATG Ggc gcc gca atg cgt cc	cta AAG CTT tca gcg cat ccg acg	68
$DosS_{11-231}$	gac gCC ATG Gac ggc gcc cga atg	c cg t AGG CTT tca ggt gcc gga caa caa ctc g	68
$DosS_{63-231}$	gta gCC ATG Gac ctg gag gca acc	c cg t AGG CTT tca ggt gcc gga caa caa ctc g	65
$DosS_{378-578}$	cttc gca aCC ATG Gca tgc gcg a	ggc tat AAG CTT cta ctg cga caa cgg tgc	70
$DosS_{COM}$	ggt gct GGA TCC cgc att cga gga ggc cag gct gcg	ctg gcg AAG CTT cgt gct ctg cca ccg cgg ctg	70
$TrcS_{1-509}$	cta gCC ATG Gcc atg atc ccg	c tct AAG CTT tca ggc ggt agt ggc	60
$MtrB_{1-568}$	cta gCC ATG Gcc atg atc ttc ggc tc	c tag AAG CTT tca acc gct cca ctc c	60
$DosT_{61-208}$	gac tag CAT ATG aag ctc gac gcc	c tag GGA TCC tca ccg tga ttc ctc	68
$DosT_{209-376}$	gac tag CAT ATG acc cgg gaa gcg tgg	c tag GGA TCC tca tcg cat ctg	68
$DosT_{COM}$	cg aat GGA TCC atg gtg gcc gcg att acc gtg	cga ata cAA GCT Tgg gtg tcg tcg acg ctg tcg	70

2.4.7 DNA ligations and transformations

2.4.7.1 Ligation of DNA fragments

Ligations were set up in 10 µl or 20 µl volumes containing 50-100 ng of vector DNA and 40-50 ng of insert DNA, 1X T4 DNA ligation buffer (NEB) (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 1 mM ATP, 25 µg/ml bovine serum albumin (BSA)) and 400 U of T4 DNA ligase (NEB). The time period and temperature for incubation of the ligations varied depending on the specific reaction conditions, but were normally left overnight at room temperature.

2.4.7.2 Transformation of *E. coli* DH5α and BL21 strains

50 µl aliquots of competent DH5α or BL21 cells were defrosted on ice; to this 5 µl of ligation mixture or 1 µl of DNA plasmid was added and left on ice for 30 minutes. The cells were then heat shocked at 42 °C for 45 seconds and transferred to ice for a further 2 minutes. 400 µl of LB broth was added to the cells and incubated at 37 °C and 200 rpm. 250 µl of these cells were plated onto agar fortified with the appropriate antibiotic (section 2.3) and incubated overnight at 37 °C.

Recombinant plasmids obtained from transformed *E. coli* DH5α cells and purified using the Qiaprep Spin Miniprep kit (Qiagen, section 2.4.1) were subjected to restriction enzyme digestion and ran on agarose gels to verify the presence of insert of correct size. Positive recombinant plasmids were then sent for sequencing (MWG, http://www.mwg-biotech.com/html/all/index.php).

2.4.8 Site directed mutagenesis

The specific point mutations H73A, H89A, H93A, H97A, H113A, H139A, and H149A were introduced into $DosS_{63-210}$ and $DosS_{63-379}$. H149A mutation was also made in $DosS_{COM}$. Stop mutations were introduced into: $DosS_{231-379}$ in place of residues: D362, A364, L366, A367, Q369, A371 and R376 and in $DosS_{231-470}$ at residues: E416 and D396. Site Directed Mutagenesis was performed using a modified version of the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene, LaJolla, CA, USA) protocol. The mutagenic oligonucleotides (Table 2-2) were synthesized (MWG) and used to perform a PCR amplification. Reactions were performed in 50 µl volumes containing 1x reaction buffer (20 mM Tris HCl pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.2 mM $MgSO_4$, 0.1 % Triton X-100,

0.1 mg mL^{-1} BSA), 150 ng forward primer, 150 ng reverse primer, 200 µM dNTP mix, 2.5 U *PfuTurbo* DNA polymerase enzyme, and 20 ng plasmid DNA.

PCR amplification was performed using the Primus thermocycler (MWG) set at 95 °C for 30 seconds before performing 16 cycles of 95 °C for 30 seconds, 55 °C for 1 minute and 68 °C for 12 minutes. 10 U of Dpn1 restriction endonuclease was added to each amplification reaction and incubated at 37 °C for 1 hour to digest parental (non-mutated) dsDNA. 10 µl aliquots of each reaction tube were run on a 1 % (w/v) agarose gel at 50 V for 1 hour to check for the presence of PCR amplified products. 5 µl of Dpn1 treated PCR reaction was used to transform 50 µl *E. coli* XL1-Blue competent cells. Transformation was done by incubating the cells and PCR products together on ice for 30 minutes before heat shocking at 42 °C for 45 seconds. The reaction tubes were further incubated on ice for 2 minutes before adding 0.5 ml of preheated (42 °C) NZY^+ broth (10 g/L casein hydrolysate, 5 g/L yeast extract,5 g/L NaCl, 20 % (w\v) glucose, 12.5 mM $MgCl_2$, 12.5 mM $MgSO_4$). The cells were incubated at 37 °C for 1 hour with shaking at 200 rpm. 250 µl of the transformation reaction was plated onto agar containing 50 µg/ml kanamycin and incubated overnight at 37 °C. Plasmid DNA was purified from single colonies and mutagenesis verified by sequencing (MWG).

2.4.9 Preparation of glycerol stocks

A single colony selected from an agar plate was inoculated into 5 ml of LB with the appropriate antibiotic. Cells were incubated overnight at 37 °C and 200 rpm. 1 ml of each culture was transferred into a cryogenic vial along with 500 µl of 70 % glycerol and stored at -80 °C.

Table 2-2: Primers used in the site directed mutagenesis experiments. In bold are the residues corresponding to the mutations.

Constructs	Sequences of Primers: 5'- 3'	
	Forward	Reverse
DosS H73A	cct acg cgc tat cgt **ggc** atc agc gac cag cct tg	caa ggc tgg tcg ctg at**g** **c**ca cga tag cgc gta gg
DosS H89A	cgc tat gga ggt **ggc** cga ccg gca gca tc	gat gct gcc ggt cg**g** **c**ca cct cca tag cg
DosS H93A	gca cga ccg gca **ggc** tcg ggt att gca ctt t	aaa gtg caa tac ccg a**gc** ctg ccg gtc gtg c
DosS H97A	gca gca tcg ggt att **ggc** ctt tgt cta tga agg cat c	gat gcc ttc ata gac aaa **ggc** caa tac ccg atg ctg c
DosS H113A	ggc gga tcg **gcg** ccc tac cga aag gcc	ggc ctt tcg gta ggg **cgc** cga tcc gcc
DosS H139A	gga cga tgt ttc tgc **ggc** acc ggc ctc gat tgg	cca atc gag gcc ggt **gcc** gca gaa aca tcg tcc
DosS H149A	ggt ttt ccg ccg tat **gct** ccg ccg atg cgt a	tac gca tcg gcg ga**g** **c**at acg gcg gaa aac c
$DosS_{231-379}$ D362Stop	atg gcc gcg ttc gcc **tag** cag gcc gcg ctg gct tg	ca agc cag cgc ggc ctg **cta** ggc gaa cgc ggc cat
$DosS_{231-379}$ A364Stop	gcg ttc gcc gac cag **tag** gcg ctg gct tgg caa ttg gcc	ggc caa ttg cca agc cag cgc **cta** ctg gtc ggc gaa cgc
$DosS_{231-379}$ L366Stop	ttc gcc gac cag gcc gcg **tag** gct tgg caa ttg gcc ac	gt ggc caa ttg cca agc **cta** cgc ggc ctg gtc ggc gaa
$DosS_{231-379}$ A367Stop	cag gcc gcg ctg **tag** tgg caa ttg gcc act tcg c	g cga agt ggc caa ttg cca **cta** cag cgc ggc ctg
$DosS_{231-379}$ Q369Stop	c cgc gct ggc ttg g**ta** att ggc cac ttc gca acg	cgt tgc gaa gtg gcc **aat** tac caa gcc agc gcg g
$DosS_{231-379}$ A371Stop	ctg gct tgg caa ttg **tag** act tcg caa cgt cgg atg cgc g	c gcg cat ccg acg ttg cga agt **cta** caa ttg cca agc cag
$DosS_{231-379}$ R376Stop	gcc act tcg caa cgt **tag** atg cgc gaa ctc gac gta ctg	cag tac gtc gag ttc gcg cat **cta** acg ttg cga agt ggc
$DosS_{231-470}$ E416Stop	gct gtc ccg cac **tag** cgt aat cct gaa gtg c	g cac ttc agg att acg **cta** gtg cgg gac agc
$DosS_{231-470}$ D396Stop	cgt gac ctc cat **tag** cat gtc atg cag cgg c	g ccg ctg gat gac atg **cta** atg gag gtc acg

2.5 Protein expression

2.5.1 Small scale protein expression tests

Small scale protein expression tests were typically performed in 100 ml LB media. A single recombinant colony selected from a plate of freshly transformed *E. coli* BL21 (DE3) pLysS or *E. Coli* Rosetta (DE3) pLysS or *E . coli* C41 (DE3) or *E. coli* C43 (DE3) cells was used to inoculate an overnight 5 ml LB culture containing antibiotic (section 2.3). The cells were grown at 37 °C and 200 rpm overnight. 1 ml of overnight culture was then used to inoculate 100 ml of LB medium fortified with antibiotic and subsequently incubated at 37 °C and 200 rpm, until the optical density (OD) of the larger culture reached OD_{600nm} = 0.6-0.8 absorbance units. The expression of protein was induced for 3 hours at 37 °C (unless otherwise stated) with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After the expression period the cells were pelleted at 6000 rpm for 15 minutes and frozen until further use.

Before and after the induction of protein expression, a 1 ml sample was taken and the OD recorded. The samples were spun down at 13,000 rpm and the pellets resuspended with water, to obtain an OD of 0.6. 10 μl of the pre- and post- induction samples were run on a polyacrylamide gel (section 2.6.6) to test for protein expression.

2.5.2 Large scale protein expression of DosS constructs

Protein expression was routinely performed in 4 L LB medium. The recombinant clones were transformed into *E. coli* BL21 (DE3) pLysS cells and selected on LB agar containing kanamycin (50 μg/ml) at 37 °C. Individual colonies or colonies from a glycerol stock were grown overnight in 10 ml of LB broth supplemented with kanamycin (50 μg/ml) at 37 °C and 200 rpm. The cultures were then subcultured into 500 ml of fresh medium supplemented with kanamycin (50 μg/ml) and grown to an OD_{600nm} of 0.4-0.6. Protein production was induced with 1 mM IPTG. After 3 hours of induction the cells were cooled to 4 °C and harvested at 6000 rpm for 15 minutes. The cell pellets were stored at -20 °C until the protein purification.

2.5.3 Large scale protein expression of isotope labelled $DosS_{231-379}$ and $DosS_{231-364}$

Large scale protein expression in minimal media was performed in 2 L quantities. The recombinant clone was transformed into *E. coli* BL21 (DE3) pLysS cells and selected

on LB agar containing kanamycin (50 μg/ml) at 37 °C. An individual colony or colonies from a glycerol stock were grown overnight in 10 ml of LB broth supplemented with kanamycin (50 μg/ml) at 37 °C and 200 rpm. 1 ml of the overnight culture was subcultured into 100 ml of minimal media supplemented with kanamycin (50 μg/ml) and left to grow for 12 hours at 37 °C and 200 rpm. Subsequently 50 ml of this culture was used to inoculate 500 ml of minimal media fortified with kanamycin (50 μg/ml) and incubated at 37 °C and 200 rpm until an OD_{600nm} value of 0.4-0.6 had been obtained. Protein production was induced with 1 mM IPTG. 3 hours after induction the cells were cooled to 4 °C and harvested at 6000 rpm for 15 minutes. The cell pellets were stored at -20 °C until protein purification.

2.6 Protein purification

2.6.1 Protein purification of the DosS constructs

All protein purification procedures were carried out at 4 °C. The cells were defrosted on ice and resuspended in buffer A (50 mM Tris pH 8.0, 250 mM NaCl, 1% Triton-X 100, 10 mM β-mercaptoehanol, 10 % glycerol, complete EDTA-free cocktail inhibitor (Roche)) and then lysed by sonication (5 x 1 min bursts with 1 min intervals). The lysate was centrifuged at 13,000 g for 1 hour. The pellet (the insoluble fraction) was resuspended in 0.2 % SDS and analysed by SDS-PAGE. The supernatant (soluble fraction) was incubated with 0.5 – 2 ml of Ni^{2+}-nitrilo-tri-acetic acid chelating beads (Ni-NTA agarose resin) (Qiagen) by rolling for 1 hour. After incubation the beads were washed four times with 15 bed volumes of buffer B (50 mM Tris pH 8.0, 250 mM NaCl, 10 mM β-mercaptoehanol, 20 mM imidazole and 10 % glycerol). The protein was eluted in buffer C (50 mM Tris pH 8.0, 250 mM NaCl, 10 mM mercaptoehanol, 200 mM imidazole and 10 % glycerol) and the eluted fractions were analysed by SDS-PAGE. His-tagged TEV protease was added in a 1:10 dilution of the total amount of protein present (section 2.6.4) and the solution left to dialyse overnight at 4 °C in buffer D (50 mM Tris pH 8.0, 250 mM NaCl, 10 mM β-mercaptoehanol and 10 % glycerol). To remove the cleaved protein tag and TEV protease the dialysed solution (if DosS $_{231-379}$) was passed over Ni-NTA agarose (Qiagen), otherwise S-protein agarose beads (Novagen) were used, the unbound cleaved protein collected and subsequently passed over Ni-NTA agarose resin (Qiagen). Proteins were buffer exchanged into either crystallisation buffer (100 mM NaCl, 50 mM Tris pH 7.5, 5 mM

BME and 5 % glycerol) or NMR buffer (20 mM sodium phosphate, pH 6.0 and 100 mM NaCl) using gel filtration.

At each stage of the purification process a 15 μl sample was removed for later analysis by polyacrylamide gel electrophoresis (section 2.6.6).

2.6.2 Purification of the membrane fractions

100ml culture of MtrB or TrcS cell pellets were defrosted on ice and resuspended in 10 ml of buffer A (50 mM Tris pH 8.0, 250 mM NaCl, 1% Triton-X 100, 10 mM β-mercaptoehanol, 10 % glycerol, complete EDTA-free cocktail inhibitor (Roche)) and subsequently lysed by sonication (5 x 1 min bursts with 1 min intervals). The lysate was centrifuged at 13,000 g for 1 hour. The supernatant was incubated on ice whilst the pellet was resuspended and homogenised in 10 ml of buffer A and centrifuged at 13,000 g for 10 minutes. The supernatant from this procedure was combined to the supernatant fraction collected previously (incubating on ice). The membrane fraction was obtained by a further ultra-centrifugation step performed at 100,000 g for 1 hour and was subsequently washed twice by resuspension in 10 ml of MEM-buffer (2 M KCl, 50 mM potassium phosphate, 5 mM EDTA and 10 % glycerol) and ultra-centrifugation at 100,000 g for 45 minutes.

The pellet from the 100,000 g spin (membrane fraction) was subsequently homogenised with MEMbuffer without KCl and analysed by SDS-PAGE (section 2.6.6).

2.6.3 Size-Exclusion Chromatography

The oligomerisation properties and protein size were estimated using either a Superdex 100, Superdex 200 or Superose 6 pre-packaged size-exclusion chromatography column and ÄKTA FPLC system (Amersham Biosciences) at 4 °C. The column was pre-equilibrated with either crystallisation buffer or NMR buffer. The protein samples (1-5 ml) were injected into the pre-equilibrated column at a flow rate of 0.5 ml/minute with 0.5-3 ml fractions collected. Eluted protein was detected using UV absorbance at 280 nm and verified by SDS-PAGE analysis (section 2.6.6).

2.6.4 Increasing the incorporation of the haem cofactor

The purification procedure was subsequently altered by the addition of 24 μM haemin to resuspended cells (cells containing the plasmids $DosS_{1-579}$, $DosS_{1-379}$, $DosS_{63-210}$ (GAFA) and $DosS_{63-379}$) prior to sonication (Bertrand *et al.*, 2004).

2.6.5 Determining protein concentration

2.6.5.1 BioRad

Protein concentrations were determined using the BioRad (BioRad) protein assay kit following the manufactures protocol. This assay is derived from the Bradford method (Bradford, 1976) and is based on the colour change of Coomassie Brilliant Blue G-250 dye, when binding to the protein residues arginine, tryptophan, tyrosine, histidine and phenylalanine. Free dye in solution has an absorption maxima at 470 and 650 nm, but when bound to protein has an absorption maximum at 595 nm. The colour change can therefore be quantified at 595 nm with a spectrophotometer and protein concentration determined according to the graphic dependence of the protein standard BSA absorbance on the BSA concentration. A set of different BSA concentrations was measured in order to draw the reference curve.

2.6.5.2 Absorbance spectra and measurement of protein concentration

UV-visible absorbance spectra of the proteins were measured in a 1 ml quartz cuvette using CARY100 Bio UV-Visible spectrometer. Molar extinction coefficients were calculated based on the amino acid sequence of the various constructs (Pace *et al.*, 1995) and are listed in Table 2-3 along with their corresponding molecular weight (Da) obtained using the program Expasy ProtParam (Gasteiger *et al.*, 2005). Protein concentrations were calculated using the Beer-Lambert Law:

$$A_{280} = \varepsilon_{280} . C . l$$

A_{280} : the absorbance of the protein at 280 nm
ϵ_{280} : the molar extinction coefficient at 280 nm ($M^{-1}cm^{-1}$)
C : the protein concentration (M)
l : the path length (cm)

Table 2-3: Summary of the Molar extinction coefficients and molecular weights for the DosS constructs.

DosS constructs	Molar Extinction Coefficient ($M^{-1}cm^{-1}$) at 280 nm	Estimated Molecular Weight (Da)
DosS $_{1-579}$	24750	62368.7
DosS $_{1-379}$	17780	40832.7
DosS $_{63-379}$	17780	34183.0
DosS $_{63-210}$	6400	16560.8
DosS $_{231-379}$	5690	15438.8
DosS $_{378-578}$	6990	21710.2

2.6.6 Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis (SDS-PAGE gel)

For separation of proteins Polyacrylamide gels were used. SDS-PAGE gels composed of a 5 % stacking layer and a 10 % or 15 % resolving layer. The components of the gel are shown in Table 2-4. Electrophoresis was carried out using the MiniProtean II gel apparatus (Biorad) connected to a PowerPac 300 power supply (Biorad). Prior to electrophoresis 5X SDS loading dye (0.5 M Tris-HCl pH 6.8, 0.5 % bromophenol blue, 10 % SDS, and 50 % glycerol) was added to each sample, heated to 95°C for 5 minutes and micro-centrifuged for 30 seconds at 13000 rpm. Unstained Precision Plus Protein standards (Biorad) were used as molecular weight markers. Polyacrylamide gels were electrophoresed at a constant voltage of 100 V through the stacking layer and 150 V through the resolving layer.

Table 2-4: Solutions for preparing one Tris-glycine SDS-Polyacrylamide electrophoresis gel, as described by Sambrook *et al.* 1989.

Solution Components	Stacking Layer Component Volumes (ml)	Resolving Layer Component Volumes (ml)	
		10 %	15 %
H_2O	1.4	1.9	1.1
30% acrylamide mix	0.33	1.7	2.5
1.0 M Tris (pH6.8)	0.25	-	-
1.5 M Tris (pH8.8)	-	1.3	1.3
10% SDS	0.02	0.05	0.05
10% ammonium persulphate	0.02	0.05	0.05
TEMED	0.002	0.002	0.002

Following electrophoresis the gel was stained in a solution containing a Coomassie dye (0.25 % (w\v) brilliant blue R Coomassie, 40 % (w/v) methanol, 7 % (v/v) acetic acid and 52 % (v/v) H_2O) and left for 15 minutes on a gel shaker. Protein bands were visualized by washing the gel in a de-staining solution (40 % (v/v) methanol, 10 % (v/v) acetic acid and 50 % (v/v) H_2O) for 15 minutes.

2.6.7 Western blot analysis

Proteins were separated on a 10 % SDS-PAGE gel and Rainbow Markers (Amersham) were used for molecular weight comparison. Gels, along with sections of blotting paper (Biorad) and a section of Polyvinylidene Difluoride (PVDF) membrane (Novagen) cut to the size of the gel, were equilibrated in western transfer buffer (Tris-buffered saline (TBS, 100 mM Tris-HCl, pH 7.5, 150 mM NaCl), 0.03 % w/v SDS, 20 % v/v methanol) for 10 minutes. Following equilibration, proteins were transferred to a PVDF membrane (Novagen) in Mini Protean II (Bio-Rad) by electroblotting in transfer buffer with a constant voltage of 25 milliamps (mA) for 16 hours. Membranes were then incubated for 30 minutes in blocking buffer (3% BSA, 0.1 % Tween 20, 1 mM EDTA) to block the remaining protein binding sites on the membrane.

The primary Mouse anti-His antibody (CloneTech) was added at a dilution of 1:750 to blocking buffer. Upon incubation for 2 hours with primary antibody, membranes were washed 5 times with TBS buffer containing 0.05 % Tween 20. The membrane was

then soaked in blocking buffer with 1:1000 peroxidase-conjugated goat anti-mouse Immunoglobulin G antibody (CloneTech), for 2 hours, and the membrane was subsequently washed three times in TBS containing 0.05 % Tween 20. Excess liquid was drained from the membranes which were analysed by exposure to Hyperfilm X-ray film (Amersham) in an autoradiography cassette (Amersham). The film was developed in an Optimax X-ray film processor (Pierce). Exposure times were 0.5 second.

2.6.8 Unfolding and refolding of $DosS_{231-379}$

^{15}N labelled GAFB $DosS_{231-379}$ was denatured by dialysing into buffer containing 20 mM sodium phosphate, 100 mM NaCl, 6 M Guanidine-HCl, pH 7.5. Denatured protein was loaded onto a Superdex 75 FPLC size-exclusion chromatography column. 2 ml protein fractions were collected, pooled and concentrated using a Vivaspin column (Millipore) to 150 µl. 150 µl of unfolded GAFB $DosS_{231-379}$ was refolded by dilution where small volumes (5 μl) were gradually added, drop by drop to 5 ml of NMR buffer (20 mM sodium phosphate, 100 mM NaCl, pH 6.0) with constant mixing.

2.7 Characterisation of purified protein

2.7.1 Circular dichroism spectroscopy

Far Ultra Violet (UV) circular dichroism (CD) analysis was carried out using an Aviv 202SF spectropolarimeter (Aviv Instruments Inc.) connected to a thermostated water bath at 25°C. CD spectrum were measured of 24 µM $DosS_{231-379}$ (100 mM sodium phosphate, pH 7.0 buffer) and 13.61 μM GAFA H149A mutant and 7.08 μM GAFA $DosS_{83-210}$ (100 mM Tris-HCl, pH 8.0, 100 mM NaCl) samples. Spectra were recorded between 190 nm and 260 nm in a 0.1 cm path length quartz cuvette, with 1 nm increments and subtracted from buffer scans. The results were represented as per-residue molar ellipticity calculated from the following equation:

$$Per\ \mathrm{Re}sidue\ Molar\ Ellipticity = \frac{\theta_{abs}}{aa.l.c.10}$$

θ_{obs} = observed ellipticity (millidegrees)

aa = Number of amino acids

l = optical path length (cm)

c = polypeptide concentration (M)

Protein secondary structure analysis was performed using the program CONTIN (Provencher and Glockner, 1981).

2.7.2 Electron Microscopy by negative staining

Electron microscopy (EM) experiments were set up in collaboration with Professor Helen Saibil at Birkbeck College, University of London with the assistance of Mr Nadav Elad. Full length DosS corresponding to a molecular mass of 550 kDa, by Superose 6 size exclusion chromatography, was used for preliminary studies by negative staining EM. The sample was diluted in a buffer containing 100 mM NaCl, 50 mM Tris pH 7.5 to give a final concentration of 400 μg/ml. 3 μl of sample were placed on a glow discharged ionized carbon coated copper grids, which were then stained with 2 % (w/v) uranyl acetate for 30 seconds. Images were viewed using a Technai T12 FEI electron microscope, operated at an accelerating voltage of 120 kV and a magnification of 42,000.

2.7.3 Nuclear Magnetic Resonance

The Nuclear Magnetic Resonance (NMR) spectra: [^{1}H,^{15}N], HSQC, three dimensional triple resonance experiments [HNCO, HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, HN(CA)CO, HA(CA)NH and HA(CACO)NH] were acquired at proton frequency of 500 or 600 MHz on the Varian Unityplus spectrometer at 25 °C (unless otherwise stated) with established pulse sequences. Parameters for each experiment are listed in Table 2-5. All NMR experiments were set up with the help of Dr R. Harris using a 0.6 mM [^{1}H, ^{13}C, ^{15}N]-labelled $DosS_{231\text{-}379}$ or $DosS_{231\text{-}364}$ protein sample. Sequence specific resonance assignments were found by correlating intra- and inter-residue resonances of CO, C_α and C_β shifts in order to create connected segments of sequence, which were then

Table 2-5: Experimental details for the 2D and 3D NMR experiments acquired for the sequential backbone assignment on [^{13}C,^{15}N] $GAFB_{231-379}$ at 298 K.

	^{1}H		^{13}C		^{15}N		
Experiment	Points	Sweep Width	Points	Sweep Width	Points	Sweep Width	Transients
HSQC[i]	512 (114ms)	4500			128 (79ms)	1620	8
HNCA[ii]	512 (114ms)	4000	64 (18ms)	3500	32 (25ms)	1300	16
HN(CO)CA[i]	512 (114ms)	4500	52 (13ms)	4000	32 (20ms)	1620	12
HNCACB[i]	512 (114ms)	4500	64 (7ms)	9200	32 (20ms)	1620	32
CBCA(CO)NH[i]	512 (114ms)	4500	64 (7ms)	9200	32 (20ms)	1620	12
HNCO[ii]	512 (114ms)	4500	40 (22ms)	1800	32 (20ms)	1620	16
HA(CA)NH[i]	512 (114ms)	4500	40 (22ms)	1800	32 (20ms)	1620	24
HA(CACO)NH[i]	512 (114ms)	4500	40 (22ms)	1800	32 (20ms)	1620	16

(i) Spectra acquired on a Varian INOVA spectrometer operating at 1H frequency of 600 MHz
(ii) Spectra acquired on a Varian UnityPlus spectrometer operating at 1H frequency of 500 MHz

assigned using the probabilistic method of Grzesiek and Bax based on the characteristic C_α and C_β shift distributions of each residue type (Grzesiek and Bax, 1993).

NMR data was processed using NMRPipe (Delaglio *et al.*, 1995) and AZARA software packages and images of 2D spectra were produced using the program PLOT2 from the AZARA package. Images of the 1D spectra were produced using the program NMRDraw.

2.7.4 Electrospray ionisation mass spectrometry

The precise molecular weight of $DosS_{231-379}$ was determined by Electrospray ionisation mass spectroscopy (ESI-MS) using a Bruker MicroTOF machine. The ESI instrument was calibrated with Tunemix positive (Agilent).

The protein sample was dialyzed into water and subsequently diluted with 50 % acetonitrile and 1% formic acid to give a final concentration of 1 mg/ml. A 50 µl sample was injected with a pump syringe at a flow rate of 200 µl/hour and separation was accomplished at a drying gas flow rate of 4 L/min with the nebulising gas set to 0.4 bar. The MS signals recorded were processed by DataAnalysis Compass 1.1 software package (Bruker).

2.7.5 Crystallisation

All purified DosS proteins were subjected to crystallisation trials. Commercially available crystallisation screens (Molecular Dimensions: structure screen I, structure screen II and MemSys; Emerald: Wizard I, Wizard II and Wizard III) were used to initially identify suitable crystallisation conditions using hanging drop vapour diffusion and under oil methods. All drops were set up for crystallisation at 20 °C, 16 °C or 4 °C using 1 µl of protein (2-10 mg mL^{-1}) and 1 µl reservoir buffer. Hanging drops were set up in 24 well plate (Hampton) using 500 µl of buffer in the well. Under oil screens were performed in 36 well plates using Als oil (Hampton).

2.7.6 Radioactive ATPase assays

Purified Kinase $DosS_{378-578}$ domain or $DosS_{1-579}$ (15 µM) was incubated with 5 µCi of [γ-^{32}P] ATP (NEN Life Sciences) in 35 µl reaction buffer containing 50 mM Tris pH 7.5, 50 mM KCl, 25 mM $MgCl_2$, 50 µM ATP at 30 °C. 10 µl of the reaction sample was taken at 15, 30 and 45 minute intervals and quenched with 3 µl of stop buffer consisting of 250 mM Tris, pH 6.8, 10 % glycerol, 1 % SDS, 280 mM β-mercaptoethanol and 0.01 %

bromophenol blue. In addition, the assays were repeated in the presence of an NO donor, DETA NONOATE. 15 µM $DosS_{1-579}$ was incubated with 5 µCi of [γ-^{32}P] ATP (NEN) and 2.5 mM diethylenetriamine-nitric oxide (DETA NONOATE, Axxora Ltd) in 35 µl reaction buffer containing 50 mM Tris pH 7.5, 50 mM KCl, 25 mM $MgCl_2$, 50 µM ATP at 30 °C. The reaction was terminated at 0, 10, 30, 60, 90 and 240 minutes by the addition of 3x SDS loading dye (250 mM Tris HCl, pH 6.8, 10 % glycerol, 1 % SDS, 280 mM β-mercaptoethanol, 0.01 % bromophenol blue). All reaction mixtures were subsequently analysed by SDS-PAGE followed by rinsing of the gel in water, drying for 2 hours and autoradiography overnight.

2.7.7 Spectrophotometric studies

All spectrophotometric experiments were set up in collaboration with Professor Peter Rich and Doug Marshall at UCL. Spectra were recorded in the range of 380-700 nm at room temperature in buffer consisting of 50 mM Tris pH 7.5, 100 mM NaCl, 5 % glycerol and 5 mM β-mercaptoethanol using a single-beam instrument built in house. An I_0 scan of 970 μl buffer in a 1 ml closed quartz cuvette was recorded as a blank before a spectrum was acquired after the addition of 30 µl 5 mM GAFA $DosS_{63-210}$ sample or $DosS_{1-579}$. Spectra were also recorded after the addition of either 1.5 μl 500 mM potassium ferricyanide, a few granules of solid sodium dithionite and addition of carbon-dioxide or nitric-oxide gas through the solution. Spectra were also acquired under the same conditions for a 5 µM bovine haemoglobin sample.

CHAPTER THREE

CLONING, EXPRESSION AND PURIFICATION OF THREE *MTB* SENSORY PROTEINS

3 Cloning, expression and purification of three *MTB* sensory proteins

The environmental stimuli to which *Mycobacterium tuberculosis* sensory proteins MtrB and TrcS respond are, as yet, unknown. DosS, on the other hand, is known to respond to hypoxia. However the molecular identity of the stimuli was not known prior to this work.

The initial objective of this project was to generate a panel of soluble DosS, MtrB and TrcS protein constructs so that suitable candidates could be identified for in-depth structural analysis and to provide a basis for probing the function and mechanism of each of the three sensory proteins. This chapter firstly describes the cloning of full length MtrB, TrcS and DosS, and truncated DosS *MTB* genes in the expression vector pET30a to obtain N-terminal hexahistidine-tagged proteins; secondly the overexpression of cloned constructs; thirdly the purification procedures using immobilised metal ion affinity chromatography and size exclusion chromatography (SEC); and finally characterisation of the solubly expressed proteins.

3.1 Gene annotation and construct design

Proteins often contain disordered regions that can serve as linkers between structurally autonomous domains and these unstructured areas can inhibit the production of soluble protein. Information from structural homologues and sequence alignments are used to design truncated constructs that contain structurally independent domains.

In previous studies TrcS, was predicted to contain two hypothetical transmembrane helices (residues 24-49 and 188-210) (Haydel *et al.*, 1999) and that full length DosS, which encodes a 579 amino acid protein contains two GAF (cyclic GMP, adenylyl cyclase, FhlA, section 1.5.2.2) domains (residues 63-205 and 231-377) and three transmembrane helices (between residues 187-203, 274-295 and 325-342) within its N-terminal portion (Saini *et al.*, 2002; Saini *et al.*, 2004). As sensing is presumably carried out by an N-terminal input region, structural sequence analysis was undertaken in order to confirm the presence of these previously predicted domains and other putative domains, as well as their location within the full length DosS, TrcS and MtrB amino acid sequences. The full-length amino acid sequences of the three proteins were submitted to the up-dated database search engines: Simple Modular Architecture Research Tool (SMART) (Schultz *et al.*, 1998), a method which uses multiple alignments to identify and annotate signalling

domain sequences; TMHMM, an algorithm for the prediction and location of TransMembrane helices, (Krogh *et al.*, 2001)); and SignalP (Nielsen *et al.*, 1997) a program which identifies the presence of a signal peptide. The structural predictions obtained from these three databases for DosS, MtrB and TrcS have been combined and are schematically represented in Figure 3-1.

3.1.1 Gene annotation for MtrB and TrcS

Each of the MtrB and TrcS proteins are predicted to contain transmembrane helices; one transmembrane segment in MtrB (residues 40-62) and two transmembrane regions in TrcS (residues 25-47 and 188-210) (Figure 3-1 (A) and (B)). Both proteins are also predicted to have a signal peptide, residues 1-38 for TrcS and residues 1-31 for MtrB. Signal peptides direct post-translational transport of a protein to the correct subcellular compartment and are then often cleaved by specialised signal peptidases. As MtrB and TrcS proteins both have putative transmembrane helices, the signal peptide may be required to direct the input domain to the extracellular side of the membrane *i.e.* the proteins may be either periplasmic-sensing or membrane-spanning histidine kinases (section 1.6.3). Residues 277-344 of TrcS and 295-362 of MtrB are predicted to contain a histidine kinase dimerisation (HisKA) domain and as expected these residues include the invariant histidine residue that is site of autophosphorylation, thought to be located at position 287 for TrcS (Haydel *et al.*, 1999) and at 305 for MtrB (Via *et al.*, 1996). Also present is an ATP binding domain (HATPase), in between residues 408-519 and 387-502 for MtrB and TrcS respectively. Both the HisKA and HATPase domains are characteristic of a conserved transmitter C-terminal kinase core (section 1.5.2.3). A hypothetical HAMP domain (found in **h**istidine kinases, **a**denylyl cyclases, **m**ethyl binding proteins and **p**hosphatases) is predicted between residues 207-269 (for TrcS) and 235-287 (for MtrB). HAMP domains are cytoplasmic linkers that are thought to play a mechanistic role in the relay of signals (section 1.5.2.4).

Since no putative domains were detected within the sensory N-terminal regions of MtrB and TrcS, only the full length proteins were cloned.

3.1.2 Gene annotation for DosS

In DosS a HisKA and a HATPase domain are predicted to encompass residues 386-452 and 488-578, respectively. The histidine residue that is autophosphorylated in the

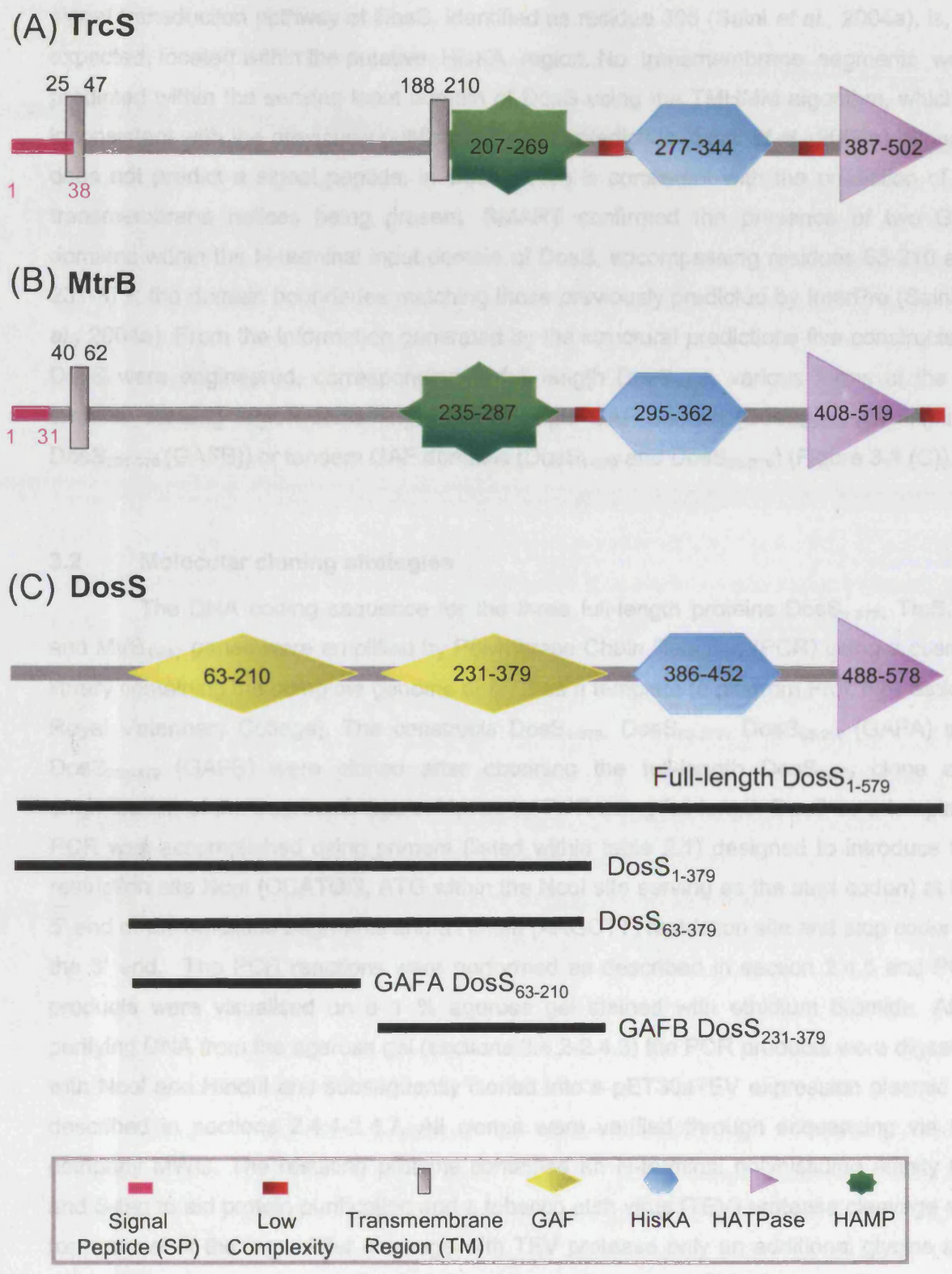


Figure 3-1: Diagram representation of the putative domains and domain boundaries predicted for TrcS (A), MtrB (B) and DosS (C). Predictions obtained using the databases SMART (Schultz *et al.*, 1998), TM-HMM (Krogh *et al.*, 2001) and SignalP (Nielsen *et al.*, 1997). All the cloned DosS variants are represented in (C).

signal transduction pathway of DosS, identified as residue 395 (Saini *et al.,* 2004a), is, as expected, located within the putative HisKA region. No transmembrane segments were predicted within the sensing input domain of DosS using the TMHMM algorithm, which is inconsistent with the previously published TMpred prediction (Saini *et al.,* 2004a). SignalP does not predict a signal peptide, in DosS which is consistent with the prediction of no transmembrane helices being present. SMART confirmed the presence of two GAF domains within the N-terminal input domain of DosS, encompassing residues 63-210 and 231-379; the domain boundaries matching those previously predicted by InterPro (Saini *et al.,* 2004a). From the information generated by the structural predictions five constructs of DosS were engineered, corresponding to full length $DosS_{1-579}$, various forms of the N-terminal sensing region containing either a single GAF domain ($DosS_{63-210}$ (GAFA) and $DosS_{231-379}$ (GAFB)) or tandem GAF domains ($DosS_{1-379}$ and $DosS_{63-379}$) (Figure 3-1 (C)).

3.2 Molecular cloning strategies

The DNA coding sequence for the three full-length proteins $DosS_{1-579}$, $TrcS_{1-509}$ and $MtrB_{1-567}$ genes were amplified by Polymerase Chain Reaction (PCR) using a cosmid library containing the complete genome of *MTB* as a template (a gift from Prof. Neil Stoker, Royal Veterinary College). The constructs $DosS_{1-379}$, $DosS_{63-379}$, $DosS_{63-210}$ (GAFA) and $DosS_{231-379}$ (GAFB) were cloned after obtaining the full-length $DosS_{1-579}$ clone and amplification of the fragments was achieved by PCR using full length DosS as a template. PCR was accomplished using primers (listed within table 2.1) designed to introduce the restriction site NcoI (CC**ATG**G, ATG within the NcoI site serving as the start codon) at the 5' end of the amplified fragments and a HindIII (AAGCTT) restriction site and stop codon at the 3' end. The PCR reactions were performed as described in section 2.4.5 and PCR products were visualised on a 1 % agarose gel stained with ethidium bromide. After purifying DNA from the agarose gel (sections 2.4.2-2.4.3) the PCR products were digested with NcoI and HindIII and subsequently cloned into a pET30aTEV expression plasmid as described in sections 2.4.4-2.4.7. All clones were verified through sequencing via the company MWG. The resulting proteins contained an N-terminal polyhistidine affinity tag and S-tag to aid protein purification and a tobacco etch virus (TEV) protease cleavage site for removal of the tags. After cleavage with TEV protease only an additional glycine and alanine residues would remain at the N-terminus of each DosS variant.

3.3 Protein overexpression

The *MTB* genes were cloned into a pET30a expression plasmid containing a kanamycin resistance gene. Within the pET expression systems, transcription of the recombinant gene is performed by the presence of bacteriophage T7 RNA polymerase, produced from an engineered *E. coli* strain such as BL21 (DE3) cells. The bacteriophage T7 RNA polymerase (T7-pol) is under control of a *lac* operon and under normal conditions of bacterial cell growth, expression of the T7-pol gene is suppressed by the binding of the *lac* repressor protein to the *lac* promoter, inhibiting transcription. The addition of allolactose (Galβ(1-6)Glc) or its non-hydrolysable analog isopropyl-β-D-thiogalactopyranoside (IPTG) to the bacterial cell culture causes the dissociation of the *lac* repressor from the *lac* promoter allowing for transcription of the recombinant gene to proceed. Since IPTG can not be broken down by *E. coli* cells the addition of this chemical to the bacterial cell culture allows for a constant, high level of heterologous protein expression.

The full-length DosS, TrcS and MtrB genes were initially cloned into *E. coli* DH5α cells, a host that does not contain the T7-pol gene, thus eliminating plasmid instability due to the production of a cloned protein potentially toxic to the host cell. For protein production, the recombinant plasmids were transformed into a BL21 (DE3) pLysS *E. coli* strain (section 2.4.7.2) containing a chromosomal copy of the gene for T7-pol. BL21 (DE3) *E. coli* cells are deficient in both *lon* and *ompT* proteases, proteolytic enzymes that can cause the hydrolysis of proteins. Small scale 100 ml Luria-Bertani (LB) cultures of each cell line were grown to an optical density (OD_{600nm}) of 0.6-0.8. The expression of the recombinant proteins was induced by the addition of IPTG to a final concentration of 1 mM (section 2.6). A sample of un-induced and induced cells were taken and diluted to an OD_{600nm} of 0.6. 10 μl of the diluted samples were subjected to SDS-PAGE analysis (section 2.6.6) to confirm protein expression.

3.3.1 Expression of *MTB* MtrB and TrcS proteins

On comparison of an un-induced and an induced sample of BL21 (DE3) pLysS *E. coli* cells containing either the TrcS or MtrB gene on a Coomassie stained SDS-PAGE gel, a small band, at the appropriate molecular weight, 62 kDa and 55 kDa respectively appear only once the cells have been induced, (Figure 3-2(A)), lanes 2 and 6). Bands to these molecular weights for both proteins are also present in the insoluble fractions (Figure 3-2(A), lanes 4 and 8), possibly confirming the presence of transmembrane regions.

Western blotting analysis, a highly sensitive method for the detection of polyhistidine-tagged proteins was used to confirm whether expression of full length MtrB

and TrcS had occurred because the levels of expression were too low to be confirmed with confidence by a Coomassie stained reducing SDS-PAGE gel. The SDS-PAGE gel containing samples from a negative control (BL21 (DE3) pLysS cells containing no DNA sequence for TrcS or MtrB), un-induced, induced, soluble fraction and insoluble fraction of MtrB and TrcS expression were transferred to polyvinylidene fluoride (PVDF) membrane and western blotting performed using an anti-HIS antibody that recognises the N-terminal hexahistidine-tag of the recombinant proteins (details of procedures can be found in section 2.6.7). The blot shown in Figure 3-2(B) confirms the expression of MtrB and TrcS *MTB* proteins, as no bands are observed in the negative control or pre-induced samples but are seen in the induced, soluble and insoluble fractions. In lanes 2 and 6, corresponding to the induced samples of MtrB and TrcS respectively, multiple bands can be seen that are also present in the insoluble fractions of MtrB and TrcS (lanes 4 and 8, respectively). The presence of multiple bands indicates that the proteins are being proteolytically cleaved within the *E. coli* cells, which could be an explanation for the poor levels of full length protein observed. The supernatant fractions of MtrB and TrcS contain solely full-length proteins (lanes 3 and 7, respectively).

The TrcS and MtrB clones were transformed into a variety of other BL21 derived expression strains, namely C41 (DE3), C43 (DE3) and Rosetta (DE3) pLysS and were tested for expression. The C41 (DE3) and C43 (DE3) strains, mutants of the BL21 (DE3) strain, have been designed for the overexpression of proteins containing transmembrane elements (Miroux and Walker, 1996), whereas the Rosetta (DE3) pLysS strain enhances the expression of proteins that contain codons rarely used in *E. coli*. A similar expression profile to the original BL21 (DE3) pLysS strain was obtained for all alternative *E. coli* strains tested.

Previous attempts to overexpress recombinant full length TrcS upon induction by the addition of IPTG, within a M15 [pREP4].pQE40 strain of *E. coli* cells, resulted in cell death (Haydel *et al.*, 1999). Cell lysis was only observed for constructs containing the two hypothetical hydrophobic membrane-spanning regions and it was concluded that the transmembrane regions within TrcS are toxic to *E. coli* (Haydel *et al.*, 1999). In contrast all of the *E. coli* strains tested survived upon induction with IPTG, *i.e.* no cell lysis was observed. In fact, these cells were able to express TrcS protein in very small quantities.

The TMHMM transmembrane algorithm, used herein, also predicts the same transmembrane elements (between residues 24-47 and 188-210) as the Kyte and Doolittle algorithm used by *Haydel et al.* (Haydel *et al.*, 1999).

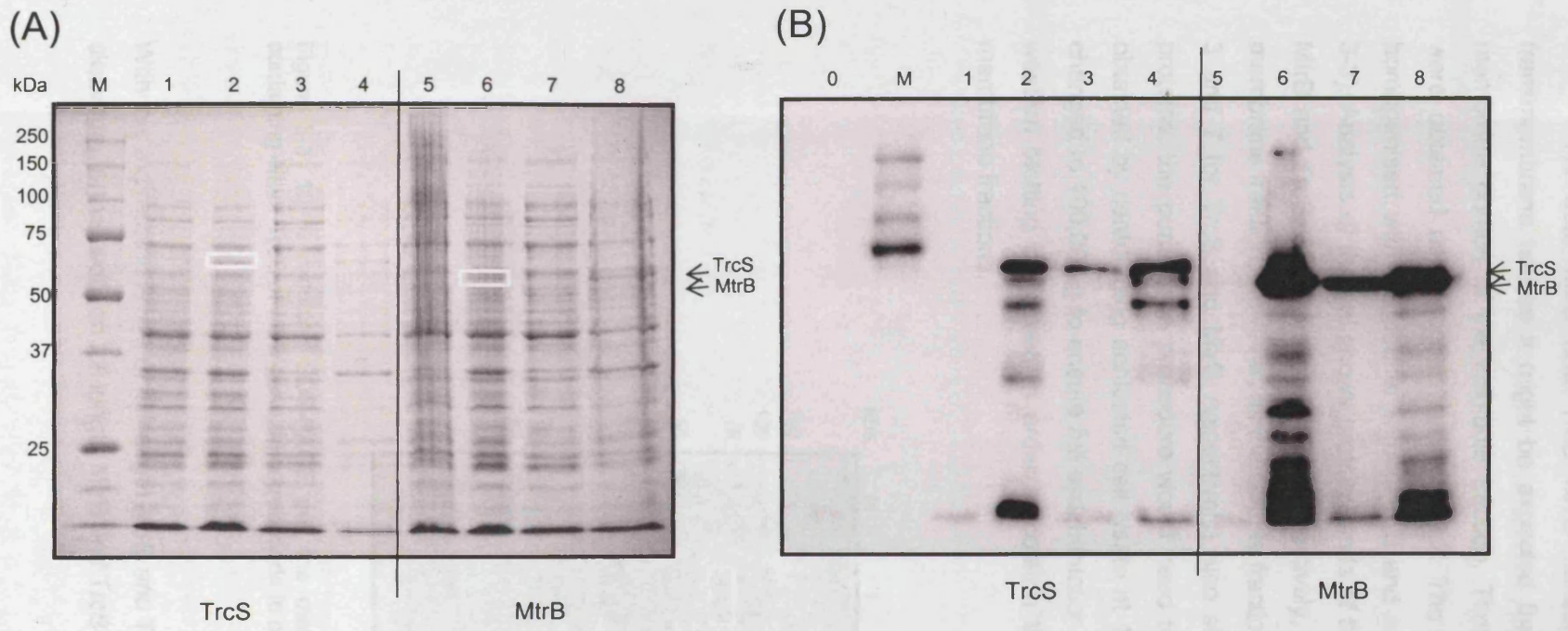


Figure 3-2: Reducing 10 % SDS-PAGE (A) and western blot (B) analysis of protein expression for MtrB (molecular weight 55 kDa) and TrcS (molecular weight 62 kDa), Lanes: M – molecular weight markers (kDa); 0 – BL21 (DE3) pLysS cells containing no DNA sequence for TrcS or MtrB (negative control); 1,5 – pre-induction; 2,6 – post induction; 3,7 – soluble fraction; 4,8 – insoluble fraction.

As both MtrB and TrcS *MTB* sensory proteins are predicted to contain transmembrane helices it might be expected that the proteins would be found in the membrane portion of the insoluble fraction. Therefore plasma membrane preparations were obtained as described section 2.6.2. The membrane fraction was subsequently homogenised with MEMbuffer without KCl and analysed on an SDS-PAGE gel (Figure 3-3). Analysis of the gel shows protein bands of the appropriate molecular mass for both MtrB and TrcS, 55 kDa and 62 kDa respectively, suggesting that the proteins are in the membrane fraction. However, as the soluble fraction for both proteins (Figure 3-2(B), lanes 3 and 7 for TrcS and MtrB respectively) also showed the presence of these sensory proteins, the purification procedure would need to be altered. The soluble fraction was obtained by centrifuging sonicated cell lysate at 13,000 g; instead this speed should be changed to 100,000 g to ensure full sedimentation of the membrane fractions. In addition, western blotting performed in order to confirm the presence of TrcS and MtrB in the membrane fractions.

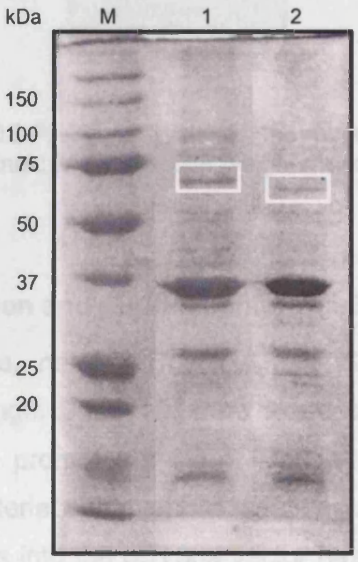


Figure 3-3: 10 % reducing SDS-PAGE gel, the membrane fractions of the C41 (DE3) cells containing either TrcS or MtrB gene, M corresponds to molecular weight markers.

With low expression levels of full length MtrB and TrcS proteins and evidence of proteolytic cleavage, further work on full length MtrB and TrcS was precluded.

3.3.2 Small scale expression of full length DosS protein

Small scale expression tests on $DosS_{1-579}$ in BL21 (DE3) pLysS *E. coli* cells clearly reveal, on a 10 % SDS-PAGE gel, an induced band (~68 kDa) that is absent in the un-induced sample (Figure 3-4). Expression tests were also repeated in Rosetta (DE3) pLysS *E. coli* cells, however, there was no difference in the amount of protein obtained.

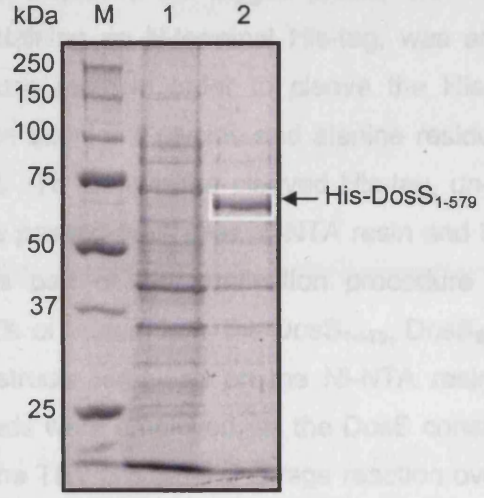


Figure 3-4: Reducing 10 % SDS-PAGE analysis of the overexpression of full length DosS, molecular weight is 62.37 kDa. Lanes: M – molecular weight markers (kDa), 1- before induction and 2 – after induction.

3.4 Large scale expression and purification of DosS constructs

Five DosS constructs, namely $DosS_{1-379}$, $DosS_{63-379}$, $DosS_{63-210}$ (GAFA) and $DosS_{231-379}$ (GAFB) and full length $DosS_{1-579}$, were expressed in BL21 (DE3) pLysS cells using the IPTG-inducible T7 promoter system. Details of the growth and induction conditions are described in material and methods, section 2.5.2.

Cloning of the proteins into the pET30a vector has the advantage of carrying the His-tag sequence, a stretch of six consecutive histidine residues that are expressed at the N-terminus of the DosS proteins. The His-tag binds to divalent cations, *e.g.* (Ni^{2+}), immobilised on a His-bind metal chelation resin (Ni-NTA resin). The nitrilo-tri-acetic acid (NTA) occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free to interact with the His-tag, which binds with high affinity and so allows for purification of the DosS proteins. Purification of all the DosS proteins was performed using Immobilised Metal Ion Affinity Chromatography and these experiments were performed using a batch method at 4 °C. The purification procedure is written in

greater detail within section 2.5 but a brief overview of the expression and purification methods are represented in Figure 3-5.

After lysis of *E. coli* cells via sonication the lysate was centrifuged and supernatant (soluble fraction) incubated with Ni-NTA agarose with agitation. The DosS protein was retained by the resin and any non-specific binding of other proteins was removed by washing with 20 mM imidazole. The DosS protein was recovered by elution with 250 mM imidazole, which displaces the tagged protein from the Ni-NTA agarose resin. TEV protease, also containing an N-terminal His-tag, was added to the solution eluted from the Ni-NTA agarose resin in order to cleave the His-tag from the DosS proteins. After cleavage only an additional glycine and alanine residues remain at the N-terminus of each DosS variant. To remove the cleaved His-tag, un-cleaved protein and TEV protease, the solution was passed back over Ni-NTA resin and the unbound cleaved protein collected. During this part of the purification procedure it was noticed that occasionally approximately 30 % of protein from the $DosS_{1-379}$, $DosS_{63-210}$ (GAFA), $DosS_{63-379}$, and full length DosS constructs remained on the Ni-NTA resin. To overcome this problem S-protein agarose beads were employed, as the DosS constructs also contained an N-terminal S-tag. Passing the TEV protease cleavage reaction over S-protein agarose beads, removed the presence of un-cleaved protein and cleaved tag; the supernatant was subsequently passed over Ni-NTA agarose to remove TEV protease. The use of S-beads prevented the loss of protein, as the presence of un-cleaved protein appeared to promote aggregation of both un-cleaved and cleaved protein onto Ni-NTA resin, possibly due to dimerisation events between the two species. The purification procedures for the DosS constructs, described above, were reproducible and no significant losses of the proteins occurred at any stages in the protocol.

In some cases a second purification stage was employed in order to remove contaminants present in small amounts. The cleaved DosS proteins were applied to a Size Exclusion Chromatography (SEC) column, a Superdex 100 for proteins $DosS_{63-379}$, $DosS_{63-210}$ (GAFA) and $DosS_{231-379}$ (GAFB) and a Superdex 200 for full length DosS protein. The different fractions eluted from the purification procedure were analysed and separated using SDS-PAGE and stained with Coomassie brilliant blue. The fractions examined for each protein were pre and post induction, soluble fraction, insoluble fraction, Ni-NTA elution, TEV protease cleavage, cleaved DosS protein and fractions collected from SEC. Typical results obtained for the purification and chromatographs obtained from SEC of $DosS_{1-379}$, $DosS_{63-210}$ (GAFA), $DosS_{231-379}$ (GAFB), $DosS_{63-379}$, and full length DosS from 4 litre cultures of *E. coli*, grown in LB media, are represented in Figure 3-6, Figure 3-7, Figure 3-8, Figure 3-9 and Figure 3-10, respectively.

The purification procedure used for all the DosS protein variants yielded proteins with >95% purity, as estimated from SDS-PAGE electrophoresis (Figures 3-4 - 3-8). The total amount of soluble protein obtained from 1 litre of *E. coli* cells for each of the DosS constructs was estimated using the BioRad protein assay kit ((Bradford, 1976), section2.6.5.1); these figures are shown in Table 3-1.

Table 3-1: Expression yields of pure protein per litre of bacterial culture grown in LB medium for the various DosS protein constructs.

Protein construct	Estimated amount of protein mg\l of *E. coli* cells
Full length DosS	4.5
$DosS_{1-379}$	4
$DosS_{63-379}$	3
GAFA	1.5
GAFB	20

GAFB protein is expressed predominately in the soluble fraction (Figure 3-8(A)), and consistently yielded approximately 20 mg/l of protein. Figure 3-8(A), lane 7 shows approximately 20 µg of GAFB protein, after the TEV cleavage reaction; and even at this concentration no visible contaminants are present. The SEC profile obtained for GAFB forms a symmetrical peak (Figure 3-8(B)), which is representative of the presence of only one protein species. Analysis of the fractions collected reveal no signs of degradation and a pure protein band of approximately 16 kDa corresponding to GAFB protein (Figure 3-8(C)). DosS GAFB protein was subsequently used in crystallisation experiments, circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy (chapter 5).

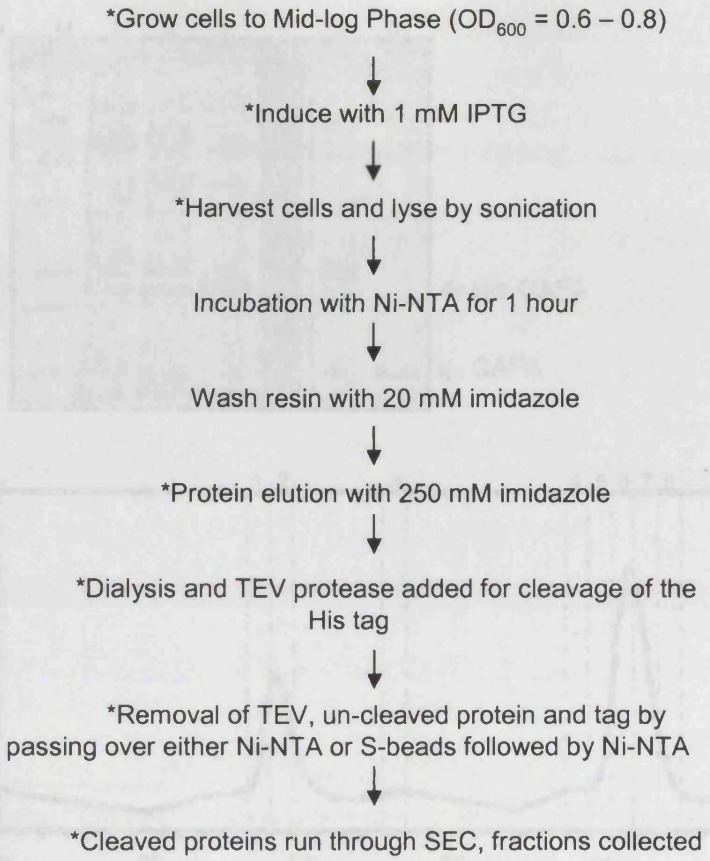


Figure 3-5: Flow diagram of the DosS purification procedure. * represent stages of the procedure analysed by SDS-PAGE.

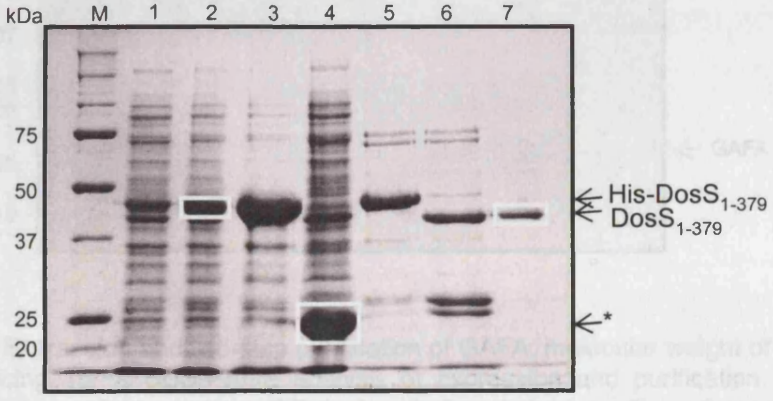


Figure 3-6: Expression and purification of $DosS_{1-379}$, molecular weight of 41 kDa.
Reducing 10 % SDS-PAGE analysis of expression and purification.
Lanes: M – molecular weight markers (kDa), 1 – before induction; 2 – after induction; 3 – insoluble fraction in 0.2 % SDS; 4 – soluble fraction; 5 – Ni-NTA agarose elution; 6 – TEV protease cleavage reaction; 7 – cleaved $DosS_{1-379}$ domain. * represents possible signs of degradation.

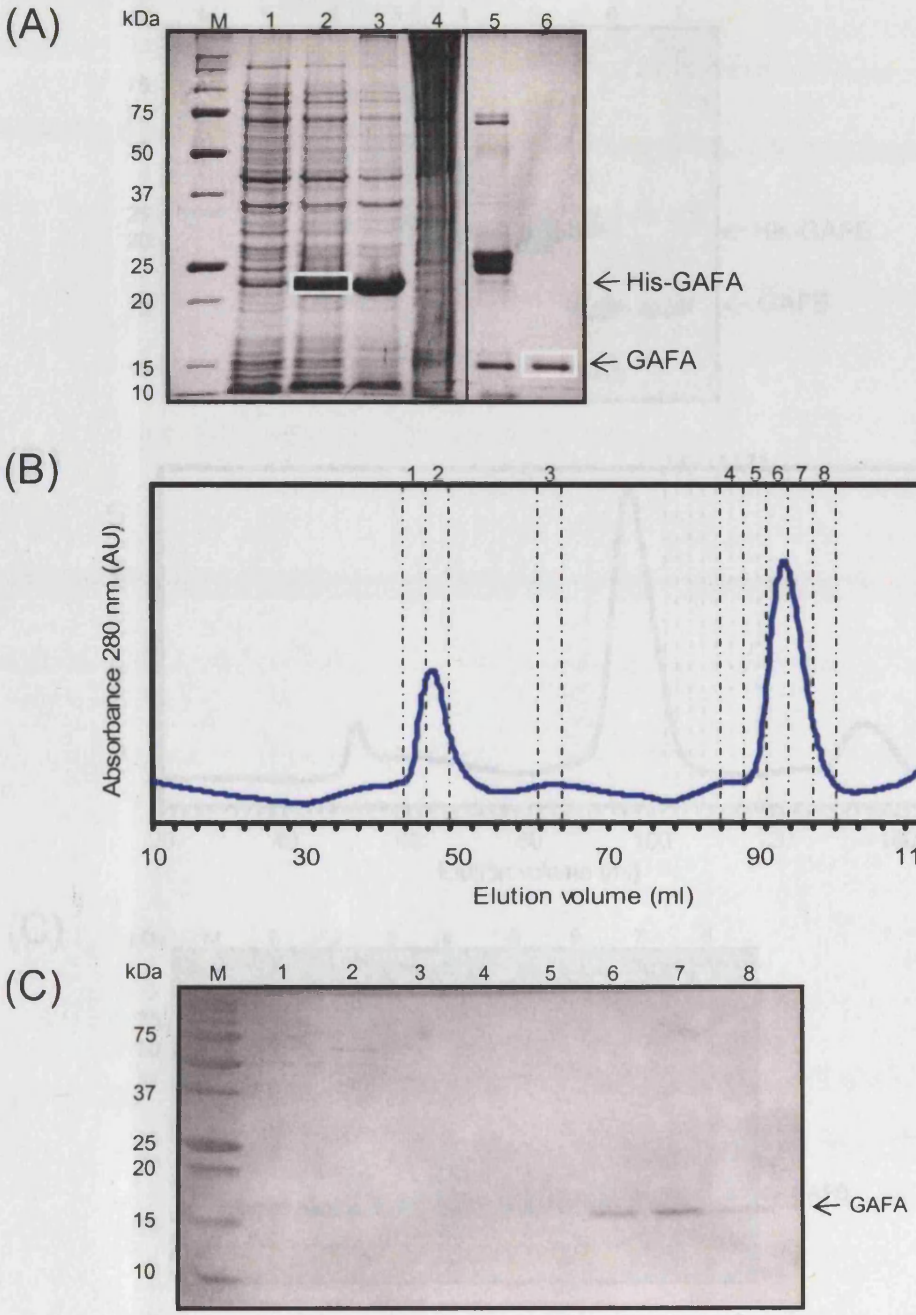


Figure 3-7: Expression and two-step purification of GAFA, molecular weight of 17 kDa.

(A) Reducing 15 % SDS-PAGE analysis of expression and purification. Lanes: M – molecular weight markers (kDa), 1 – before induction; 2 – after induction; 3 – insoluble fraction in 0.2% SDS; 4 – soluble fraction; 5 – TEV protease cleavage reaction; 6 – Ni-NTA flow through after TEV cleavage: cleaved GAFA.

(B) Size exclusion chromatography (SEC) using a Superdex 100 column, fractions collected are marked above the chromatogram.

(C) Reducing 15 % SDS-PAGE analysis of fractions from SEC (panel B). Lanes: M – molecular weight markers (kDa), other lane numbers correspond with fraction number from SEC. Fraction 6-8 were combined and concentrated for used in crystallisation trials (section 3.5.3).

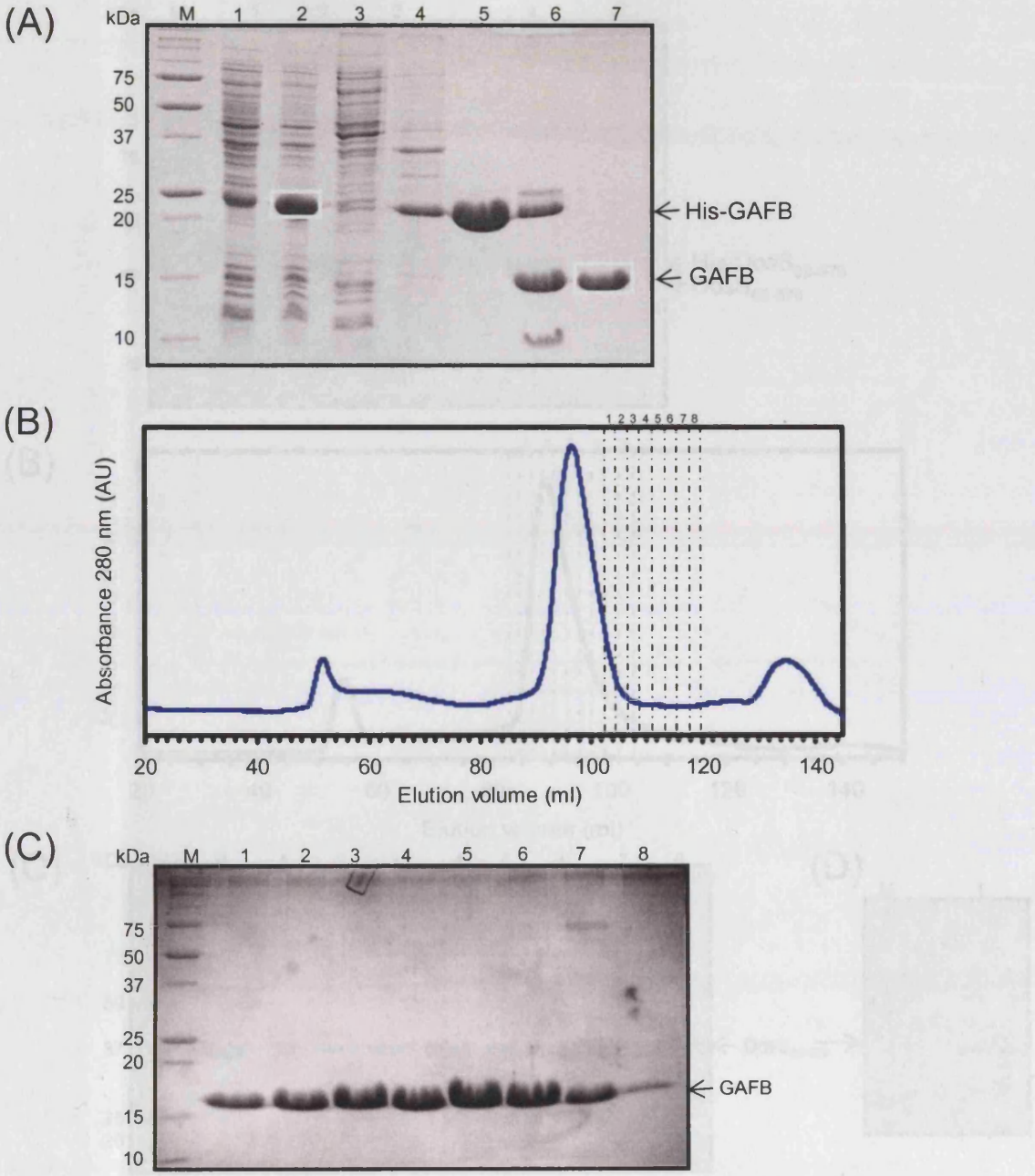


Figure 3-8: Expression and two-step purification of GAFB, molecular weight of 16 kDa.

(A) Reducing 15 % SDS-PAGE analysis of expression and purification. Lanes: M – molecular weight markers (kDa), 1 – before induction; 2 – after induction; 3 – insoluble fraction in 0.2 % SDS; 4 – soluble fraction; 5 – Ni-NTA agarose elution; 6 – TEV protease cleavage reaction; 7 – Ni-NTA flow through after TEV cleavage: cleaved GAFB.

(B) Size exclusion chromatography (SEC) using a Superdex 100 column, fractions collected are marked above the chromatogram.

(C) Reducing 15 % SDS-PAGE analysis of fractions from SEC (panel B). Lanes: M – molecular weight markers (kDa), other lane numbers correspond with fraction number from SEC. Fractions 1-6 were subsequently combined and concentrated for use in crystallisation trials (section 3.5.3).

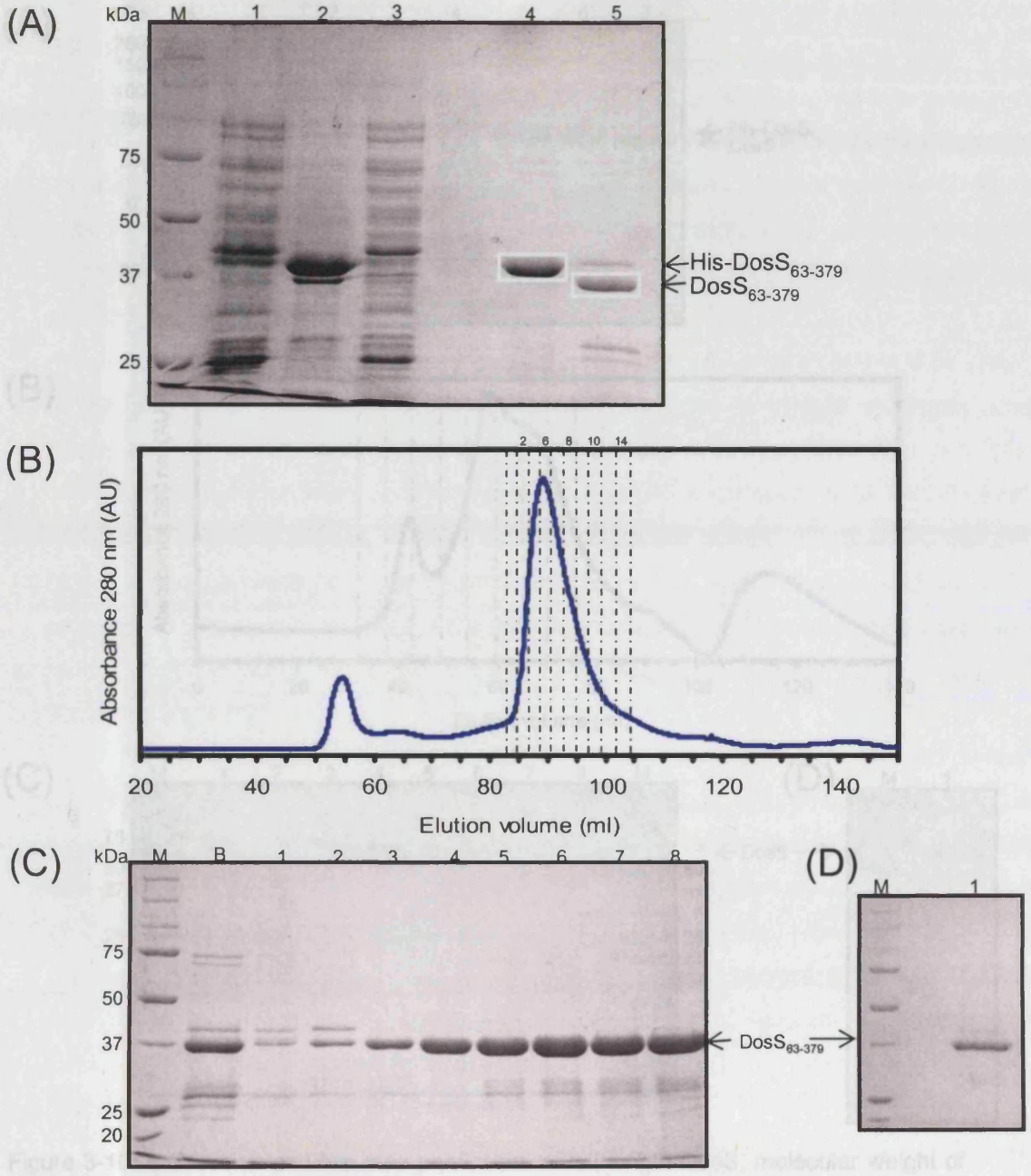


Figure 3-9: Expression and two-step purification of $DosS_{63-379}$, molecular weight of 34 kDa.

(A) Reducing 10 % SDS-PAGE analysis of expression and purification. Lanes: M – molecular weight markers (kDa), 1 – soluble fraction; 2 – insoluble fraction in 0.2 % SDS; 3 – flow through from Ni-NTA resin; 4 – Ni-NTA agarose elution; 5 – TEV protease cleavage reaction.

(B) Size exclusion chromatography (SEC) using a Superdex 100 column, fractions collected are marked above the chromatogram.

(C) Reducing 10 % SDS-PAGE analysis of fractions from SEC (panel B). Lanes: M – molecular weight markers (kDa), other lane numbers correspond with fraction number from SEC.

(D) Fractions 3-8 from SEC were combined and passed over S-beads and Ni-NTA resin. This sample was concentrated and the $DosS_{63-379}$ protein used in crystallisation trays.

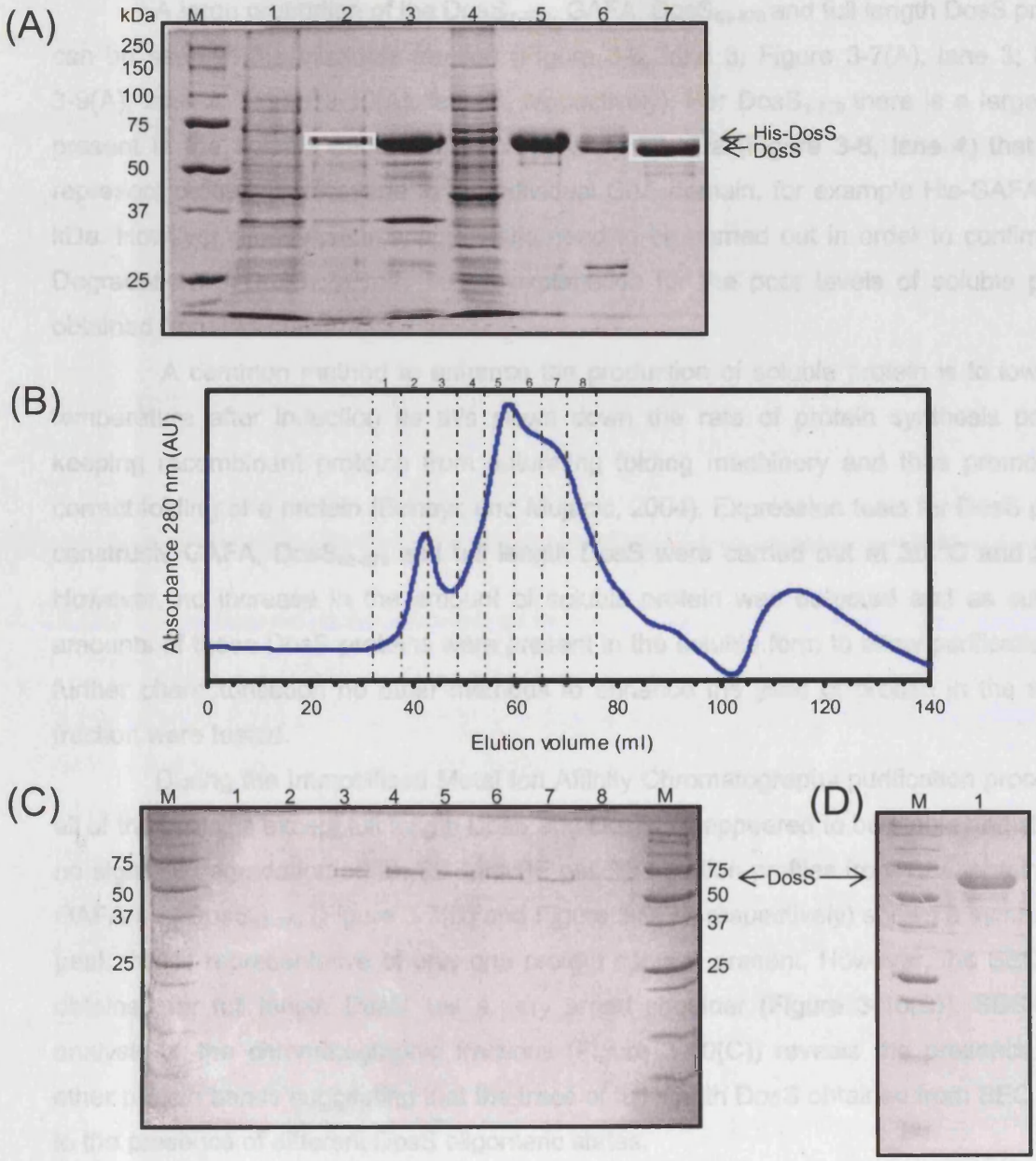


Figure 3-10: Expression and two-step purification of full length DosS, molecular weight of 62 kDa.

(A) Reducing 10 % SDS-PAGE analysis of expression and purification. Lanes: M – molecular weight markers (kDa), 1- before induction; 2 – after induction; 3 – insoluble fraction in 0.2 % SDS; 4 – soluble fraction; 5 – Ni-NTA agarose elution; 6 – TEV protease cleavage reaction; 7 – Ni-NTA flow through after TEV cleavage: cleaved full length DosS.

(B) Size exclusion chromatography (SEC) using a Superdex 200 column, fractions collected are marked above the chromatogram.

(C) Reducing 10 % SDS-PAGE analysis of fractions from SEC (panel B). Lanes: M – molecular weight markers (kDa), other lane numbers correspond with fraction number from SEC.

(D) Reducing 10 % SDS-PAGE, Lanes: M – molecular weight markers (kDa); 1 – full length DosS protein combined using fractions 4-7 from SEC and concentrated.

A large proportion of the $DosS_{1-379}$, GAFA, $DosS_{63-379}$ and full length DosS proteins can be seen in the insoluble fraction (Figure 3-6, lane 3; Figure 3-7(A), lane 3; Figure 3-9(A), lane 2; Figure 3-10(A), lane 3, respectively). For $DosS_{1-379}$ there is a large band present in the soluble fraction at approximately 23 kDa (Figure 3-6, lane 4) that could represent proteolytic cleavage to an individual GAF domain, for example His-GAFA is 22 kDa. However mass spectroscopy would need to be carried out in order to confirm this. Degradation of $DosS_{1-379}$ may be an explanation for the poor levels of soluble protein obtained from this construct.

A common method to enhance the production of soluble protein is to lower the temperature after induction as this slows down the rate of protein synthesis possibly keeping recombinant proteins from saturating folding machinery and thus promote the correct folding of a protein (Baneyx and Mujacic, 2004). Expression tests for DosS protein constructs GAFA, $DosS_{63-379}$ and full length DosS were carried out at 30 °C and 24 °C. However, no increase in the amount of soluble protein was detected and as sufficient amounts of these DosS proteins were present in the soluble form to allow purification and further characterisation no other methods to enhance the yield of protein in the soluble fraction were tested.

During the Immobilised Metal Ion Affinity Chromatography purification procedure, all of the proteins except full length DosS and $DosS_{1-379}$ appeared to be stable and showed no signs of degradation on an SDS-PAGE gel. The elution profiles from SEC obtained for GAFA and $DosS_{63-379}$ (Figure 3-7(B) and Figure 3-9(B), respectively) shows a symmetrical peak that is representative of only one protein species present. However, the SEC trace obtained for full length DosS has a very broad shoulder (Figure 3-10(B)). SDS-PAGE analysis of the chromatographic fractions (Figure 3-10(C)) reveals the presence of no other protein bands suggesting that the trace of full length DosS obtained from SEC is due to the presence of different DosS oligomeric states.

During the purification procedure it was noticed that four out of the five protein constructs (all except GAFB) had an associated light-brown colour that remained even after further purification by using SEC. Figure 3-11 shows, to illustrate the colour difference, a picture of the concentrated solutions of GAFA and GAFB (both at ~2 mg/ml). Presence of a colour would indicate a protein-associated chromophore. Exploration of the type of ligand bound to DosS is described in Results Chapter 4: Ligand binding; implications for DosS.

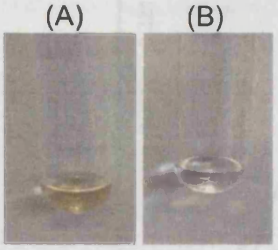


Figure 3-11: Solutions of purified protein expressed and purified by immobilised metal ion affinity chromatography and SEC (A) GAFA (B) GAFB.

3.5 Structural characterisation

3.5.1 Size Exclusion Chromatography

SEC was utilised in the purification procedure of the DosS variants to remove protein contaminants but can also be used to determine the native size of a protein and an estimate of its oligomeric state. SEC separates protein molecules according to size. As the protein solution passes through a column packed with porous polymer beads smaller molecules enter these pores and become trapped while larger molecules pass more quickly between the beads and thus elute off the column first. The molecular mass of a protein can therefore be estimated by comparing its elution volume to the elution volume of a set of known molecular weight standards that have been applied to the column. The Superdex 100 SEC column was calibrated with Blue Dextran (2000 kDa), Albumin (67 kDa), Ovalbumin (43 kDa), Chymotrypsin (25 kDa) and Ribonuclease (13.7 kDa) (Figure 3-12(A), (B)). The Superdex 200 column was calibrated with Thyroglobulin (670 kDa), Gamma Globulin (158 kDa), Ovalbumin (43 kDa), Myoglobin (17 kDa) and Vitamin B-12 (1.35 kDa) (Figure 3-12(C), (D)). The elution volume of Blue Dextran and Thyroglobulin was used to obtain the void volume for each column *i.e.* at this elution the molecules are above the upper fractionation limit. The elution volume (ml) of each protein standard (except Blue Dextran or Thyroglobulin) was plotted against the logarithm_{10} of its molecular weight (MW) generating the equation: $y = -0.0263x + 4.068$ for the Superdex 100 column (Figure 3-12(B)) and $y = -0.0619x + 9.3214$ for the Superdex 200 column (Figure 3-12(D)), where y is $\text{logarithm}_{10}(MW)$ and x the elution volume. These calibrations were subsequently used to calculate the approximate molecular weight for each of the DosS

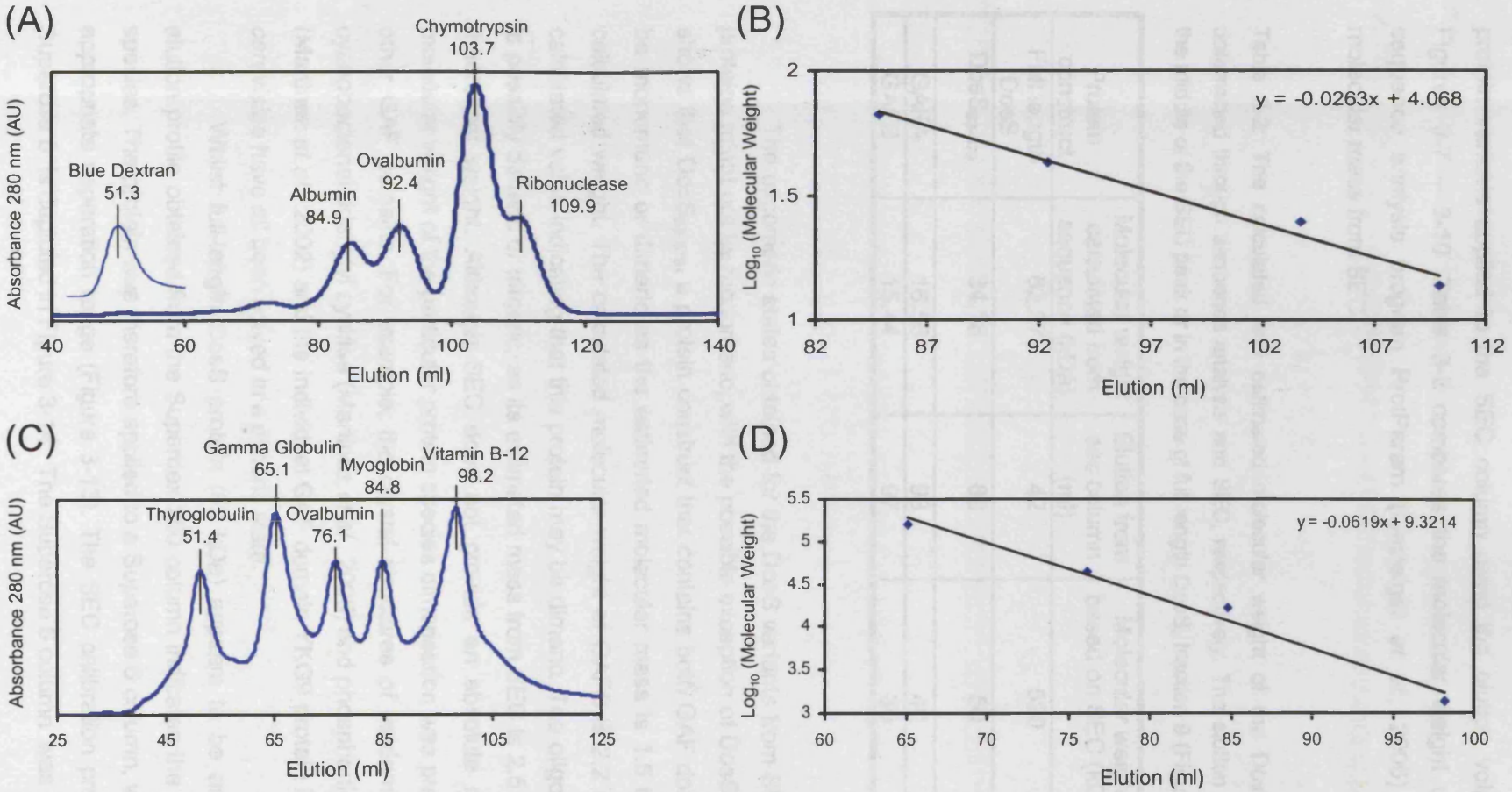


Figure 3-12: Calibration curves for the Superdex 100 and Superdex 200 columns.

(A) Protein standards used in Size Exclusion Chromatography (SEC) to calibrate a Superdex 100 column: Blue Dextran (2000 kDa), Albumin (67 kDa), Ovalbumin (43 kDa), Chymotrpsin (25 kDa) and Ribonuclease (13.7 kDa). The elution volume for each standard is marked upon the graph.

(B) Protein standards used in Size Exclusion Chromatography (SEC) to calibrate a Superdex 200 column: Thyroglobulin (670 kDa), Gamma Globulin (158 kDa), Ovalbumin (43 kDa), Myoglobin (17 kDa) and Vitamin B-12 (1.35 kDa). The elution volume for each standard is marked upon the graph.

(C)(D) Linear fit of the logarithm of the molecular weight versus the elution volume. The apparent molecular weights of DosS proteins is calculated using the equation of a straight line.

protein variants applied to the SEC column using the elution volumes obtained from Figures 3-7 – 3-10. Table 3-2 compares the molecular weight calculated using the sequence analysis program ProtParam (Gasteiger *et al.*, 2005) and the estimated molecular mass from SEC.

Table 3-2: The calculated and estimated molecular weight of the DosS protein variants as determined through sequence analysis and SEC, respectively. The elution values are taken from the middle of the SEC peak or in the case of full length DosS, fraction 9 (Figures 3-7 – 3-10).

Protein construct	Molecular weight calculated from sequence (kDa)	Elution from sec column (ml)	Molecular weight based on SEC (kDa)	Oligomerisation state
Full length DosS	62.37	42	520	Octamer
$DosS_{63-379}$	34.18	89	50	monomer / dimer
GAFA	16.56	93	40	dimer / trimer
GAFB	15.44	97	30	dimer

The oligomeric states obtained for the DosS variants from SEC indicates that the proteins might not be monomeric, with the possible exception of $DosS_{63-379}$. The SEC data shows that $DosS_{63-379}$, a protein construct that contains both GAF domains could possibly be monomeric or dimeric as the estimated molecular mass is 1.5 times higher than its calculated weight. The calculated molecular weight of GAFB is 2.2 times higher than its calculated value indicating that this protein may be dimeric. The oligomeric state of GAFA is possibly dimeric or trimeric, as its estimated mass from SEC is 2.5 times higher than its theoretical weight. Although SEC does not provide an absolute measurement of the molecular weight of the particular protein species dimerisation was previously observed for other GAF domains. For example, the crystal structures of tandem GAF domains from cyanobacterial adenylyl cyclase (Martinez *et al.*, 2005) and phosphodiesterase 2A (mouse) (Martinez *et al.*, 2002) and the individual GAF domain YKG9 protein from *Saccharomyces cerevisiae* have all been solved in a dimeric state.

Whilst full-length DosS protein (62 kDa) appears to be an octamer, the SEC elution profile obtained from the Superdex 200 column indicates the presence of multiple species. The protein was therefore applied to a Superose 6 column, which has more of an appropriate separation range (Figure 3-13). The SEC calibration profile and plot for the Superose 6 is depicted in Figure 3-14. The Superose 6 column was able to resolve the

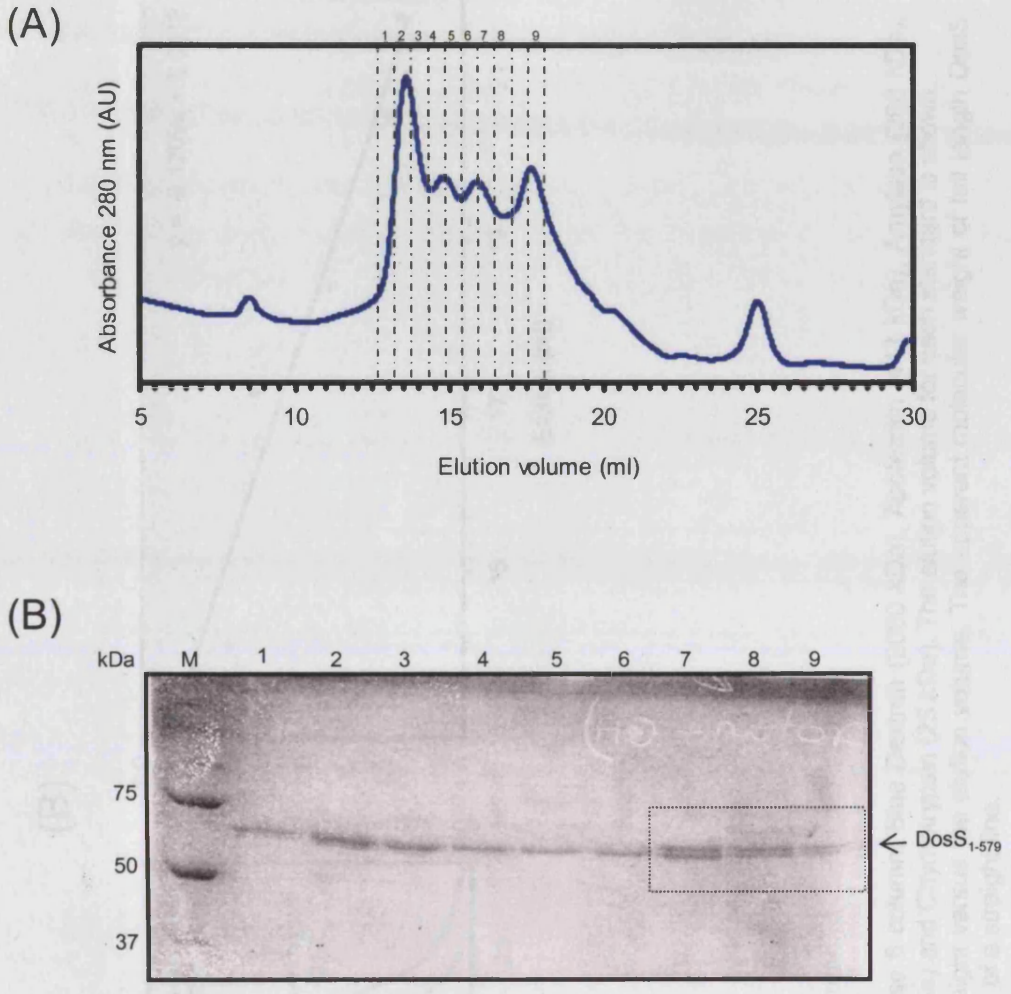


Figure 3-13: SEC of full length DosS, molecular weight of 62 kDa.

(A) Size exclusion chromatography (SEC) using a Superose 6 column, fractions collected are marked above the chromatogram.

(B) Reducing 10 % SDS-PAGE analysis of fractions from SEC. Lanes: M – molecular weight markers (kDa), other lane numbers correspond with fraction number from SEC. Two protein bands can be seen in lanes 7-9.

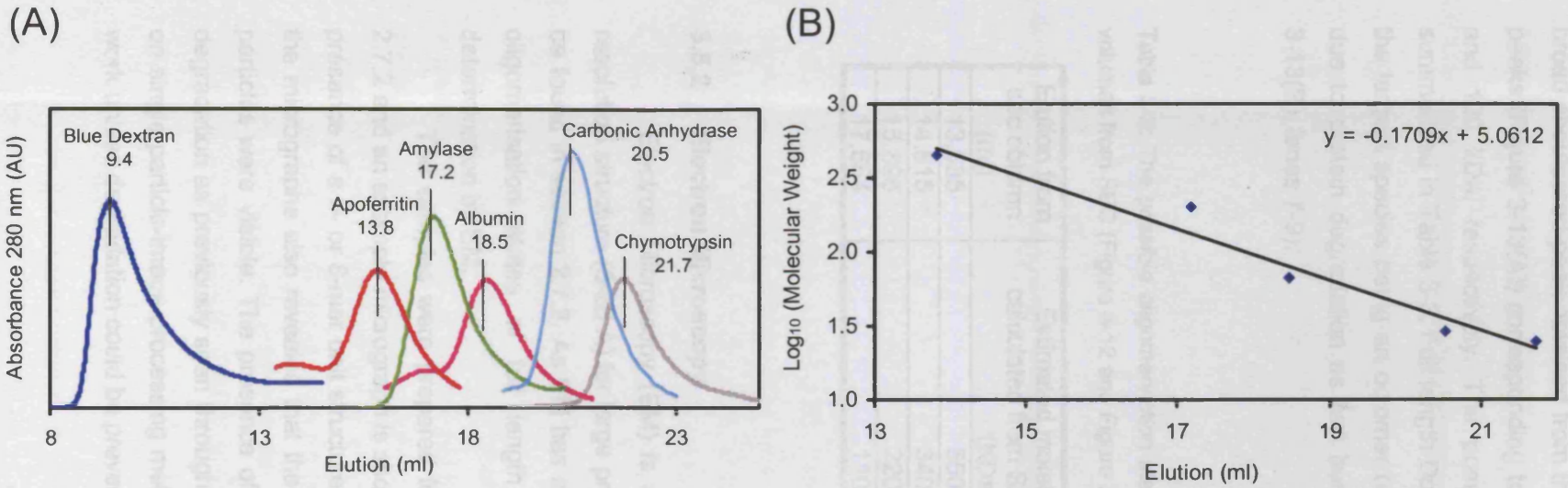


Figure 3-14: Calibration curve for the Superose 6 column.

(A) Protein standards used to calibrate a Superose 6 column: Blue Dextran (2000 kDa), Apoferritin (443 kDa), Amylase (200 kDa), Albumin (67 kDa), Carbonic Anhydrase (29 kDa) and Chymotrypsin (25 kDa). The elution volume for each standard is shown.

(B) Linear fit of the logarithm of the molecular weight versus the elution volume. The apparent molecular weight of full length DosS protein was then calculated using the equation of a straight line.

broad shouldered-peak obtained from the Superdex 200 (Figure 3-10(B)) into four distinct peaks (Figure 3-13(A)) corresponding to approximate molecular masses of 550, 350, 200 and 100 kDa, respectively. The possible oligomeric states of full length DosS are summarised in Table 3-3. Full length DosS appears to form multiple oligomeric states with the largest species being an octomer or nonomer. The smaller oligomeric states may be due to protein degradation as two bands are observed on an SDS-PAGE gel (Figure 3-13(B), lanes 7-9).

Table 3-3: The possible oligomerisation states of full length DosS as calculated using the elution volumes from SEC (Figure 3-12 and Figure 3-13).

Elution from sec column (ml)	Estimated molecular weight, calculated from SEC calibration (KDa)	No. of times higher than the calculated molecular mass
13.585	550	8-9
14.815	340	5
15.895	220	3-4
17.664	110	1-2

3.5.2 Electron Microscopy

Electron Microscopy (EM) is a technique that can be used to determine low resolution structure (5-20 Å) for large proteins or complexes (>150 kDa), more details can be found in section 2.7.2. As EM has a lower molecular size limit of 150 kDa, the higher oligomerisation states of full length DosS are potentially amenable to structure determination by EM.

The samples were prepared for negative stain EM as described within section 2.7.2 and an example micrograph is shown in Figure 3-15. The scans taken did reveal the presence of a 9- or 8-mer unit structure consistent with the SEC data obtained. However, the micrographs also revealed that the sample was not homogeneous, as various sized particles were visible. The presence of a heterogeneous sample may be due to protein degradation as previously seen through SDS-PAGE analysis. As negative stain EM relies on single-particle-image-processing methods the protein was not amenable to further EM work unless degradation could be prevented and a homogenous sample prepared.

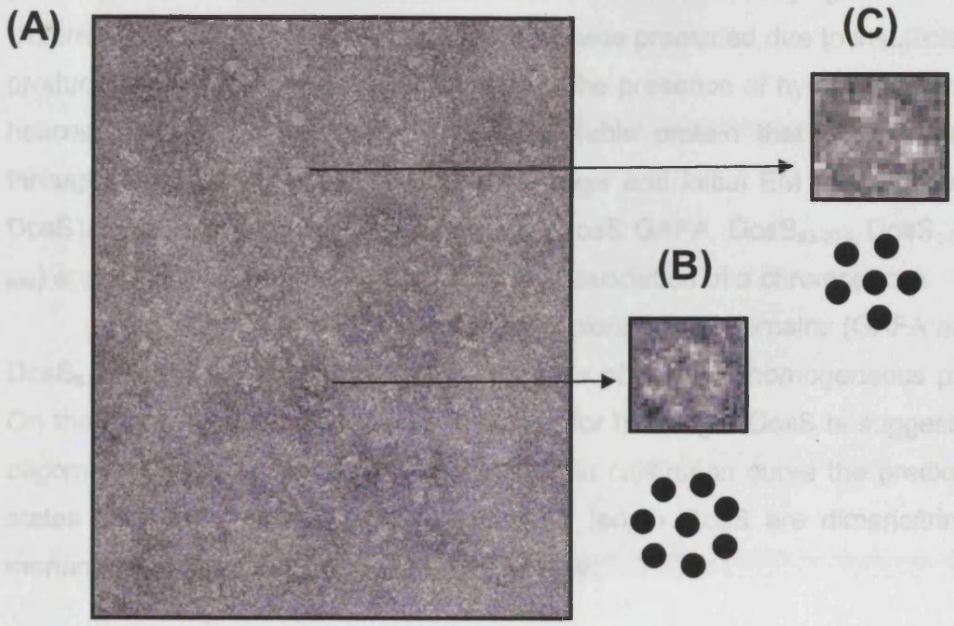


Figure 3-15: Electron micrograph of full length DosS (A) A portion of a micrograph showing particles of full length DosS protein (100 mM NaCl, 50 mM Tris-HCl, pH 7.5) which were stained with 2 % uranyl actate on carbon coated grids. Photographs were taken with a 120 kV transition electron microscope and the developed films were digitised by scanning with a resolution of 14 μm/pixel. Small particles can be seen of various sizes; the majority are composed of two layers containing a 9-mer or 8-mer unit structure (B), (C) shows a 6-mer unit.

3.5.3 Crystallisation trials

Varying concentrations of proteins GAFA, GAFB and $DosS_{63-379}$ (2-10mg/ml) were subjected to a series of sparse matrix crystallisation screens (Molecular Dimensions: structure screen I, structure screen II and MemSys; Emerald: Wizard I, Wizard II and Wizard III), using both hanging drop and micro-batch under oil methods (as described in method section 2.7.5). Trays were set up at 4°C, 16°C and 20°C. No hits were obtained for GAFA or $DosS_{63-379}$ but small needle crystals of 5 mg/ml GAFB were observed; this is discussed further in results chapter 5: Structural investigations of DosS GAFB.

3.6 Summary

The DNA constructs of full length *MTB* proteins TrcS, MtrB, DosS, DosS GAFA, DosS GAFB, $DosS_{63-379}$ and $DosS_{1-379}$ were cloned into a pET30a expression vector. All

proteins were expressed in *E. coli* cells and showed varying levels of expression. Unfortunately, further work on TrcS and MtrB was precluded due to insufficient amounts of produced full length protein, possibly due to the presence of hypothetical transmembrane helices. The DosS constructs produced soluble protein that allowed further analysis through SEC, setting up of crystallisation trays and initial EM investigations (full length DosS). Four of the purified DosS proteins (DosS GAFA, $DosS_{63-379}$, $DosS_{1-379}$ and $DosS_{1-579}$) appeared brown in colour, indicating the association of a chromophore.

The SEC profiles for the individually cloned GAF domains (GAFA and GAFB) and $DosS_{63-379}$ show a symmetrical peak indicative of a single homogeneous protein sample. On the other hand the SEC profile obtained for full length DosS is suggestive of multiple oligomeric states. Based on a globular protein calibration curve the predicted oligomeric states for GAFA, GAFB, $DosS_{63-379}$ and full length DosS are dimeric/trimeric, dimeric, monomeric/dimeric and octameric, respectively.

CHAPTER FOUR

LIGAND BINDING; IMPLICATIONS FOR DosS

4 Ligand binding; implications for DosS function

DosS is the only two component sensor for which stimuli have been identified; DosS/DosR is known to be involved in the genetic response of *MTB* to hypoxia and NO (Kendall *et al.*, 2004; Sherman *et al.*, 2001; Voskuil *et al.*, 2003), although the precise molecular basis of this response is still to be unravelled.

During the immobilised metal ion affinity chromatography (IMAC) purification procedure it was noticed that four out of the five DosS constructs (full-length $DosS_{1-579}$, $DosS_{1-379}$, $DosS_{63-379}$, GAFA $DosS_{63-210}$ surprisingly had an associated light brown-orange colour. After further purification using size-exclusion chromatography all four of the protein solutions were still coloured suggesting the association of a chromophore. As all four proteins, through SDS-PAGE analysis, were shown to be purified to homogeneity, the co-factor must be associated with the DosS protein rather than binding a contaminant or another protein that binds to DosS.

This chapter describes identification of the co-factor associated with DosS and investigation of its role in the function of DosS.

4.1 Identification of the chromophore bound to DosS

The brown-orange chromophore bound to DosS absorbs light in the ultraviolet and visible (UV-VIS, 200-800 nm) region of the spectrum. The analysis of the UV-VIS spectrum acquired on the protein can be used to identify the type of co-factor bound, as each chemically-distinct ligand gives rise to a unique spectrum. For example, an oxidised flavoprotein with flavin mononucleotide as a native cofactor, gives a characteristic UV-VIS spectrum containing absorption maxima at approximately 455, 370 and 272 nm (Wasserfallen *et al.*, 1995).

The energy absorbed by chemical groups from UV-VIS radiation causes an excitation of an electron to a higher energy state. The amount of energy required to excite the electron can be plotted against the wavelength to obtain a UV-VIS spectrum. The sample within a quartz cuvette is placed into the spectrophotometer. A UV-VIS beam in the 200-800 nm range is filtered through a monochromator that produces a single wavelength of incident radiation (I_0) at a time that is passed through the sample and the intensity of the radiation transmitted by the sample (I_1) is recorded for each wavelength. Transmittance (T) is defined as the percentage of I_0 that is transmitted through a sample

($T = (I_1/ I_0) \times 100$) and is used to calculate the amount of radiation absorbed (A) ($A = \log_{10} (I_0 / I_1)$) (Thomas, 1997).

UV-VIS spectra of the purified DosS protein constructs, full-length DosS, $DosS_{1-379}$, $DosS_{63-379}$, GAFA and GAFB were recorded in order to identify the type of cofactor present within DosS. All DosS variants except GAFB exhibited absorption maxima at 350 nm, 424 nm, 541 nm, 571 nm and at 280 nm (280 nm arises from the tyrosine and tryptophan residues of the protein). The UV-VIS spectrum of $DosS_{1-379}$ is shown in Figure 4-1(A), and reveals the characteristic spectral properties corresponding to the presence of a haem prosthetic group (Figure 4-1(B)). The absorption maxima at 571 nm, 541 nm, 424 nm and 350 nm are known as the alpha, beta, gamma (also called Soret) and delta bands respectively (Gilles-Gonzalez *et al.,* 1991; Gilles-Gonzalez and Gonzalez, 2005). Since identical Visible spectra were observed for all of the DosS protein constructs except GAFB $DosS_{231-379}$, which lacks the first GAF domain, it was concluded that the haem molecule must be binding to the shortest construct, the proximal GAF domain, GAFA. This thesis reports for the first time not only that the *MTB* DosS sensory protein binds haem but that a GAF domain is able to bind haem.

4.2 Haem-binding proteins (Haemoproteins)

Haem-binding proteins or Haemoproteins, which are ubiquitously distributed in nature, contain a haem prosthetic group derived from porphyrin with a centrally bound iron atom. These proteins are involved in a diverse range of biological functions that include the binding and transport of oxygen (*e.g.* myoglogin and haemoglobin), respiratory and photosynthetic electron transport pathways (*e.g.* cytochromes), deactivation of reactive oxygen compounds (*e.g.* catalase) and the sensing of diatomic gases (oxygen sensor *Bradyrhizobium* FixL protein and nitric oxide sensor SONO from *Clostridium botulinum*).

Chemical substitutions at the 2, 4, 5, 6 and 8 positions of a porphyrin ring give rise to four main classes of haem, known as A, B, C and D (Figure 4-2). Each class of haem can be distinguished from the other classes by its unique spectral characteristics, especially in the Visible region (350 nm to 700 nm). The most common class of haem is type B which has a protoporphyrin IX ring, as found within haemoglobin. The B type haem contains a planar porphyrin derivative consisting of four pyrrole rings linked by methine bridges, with a centrally bound iron atom. As with haem-binding PAS domains (structurally similar to GAF domains), such as FixL a rhizobial oxygen sensor, the iron atom has six coordination sites: four that are linked to the nitrogen atoms of the pyrrole ring, the

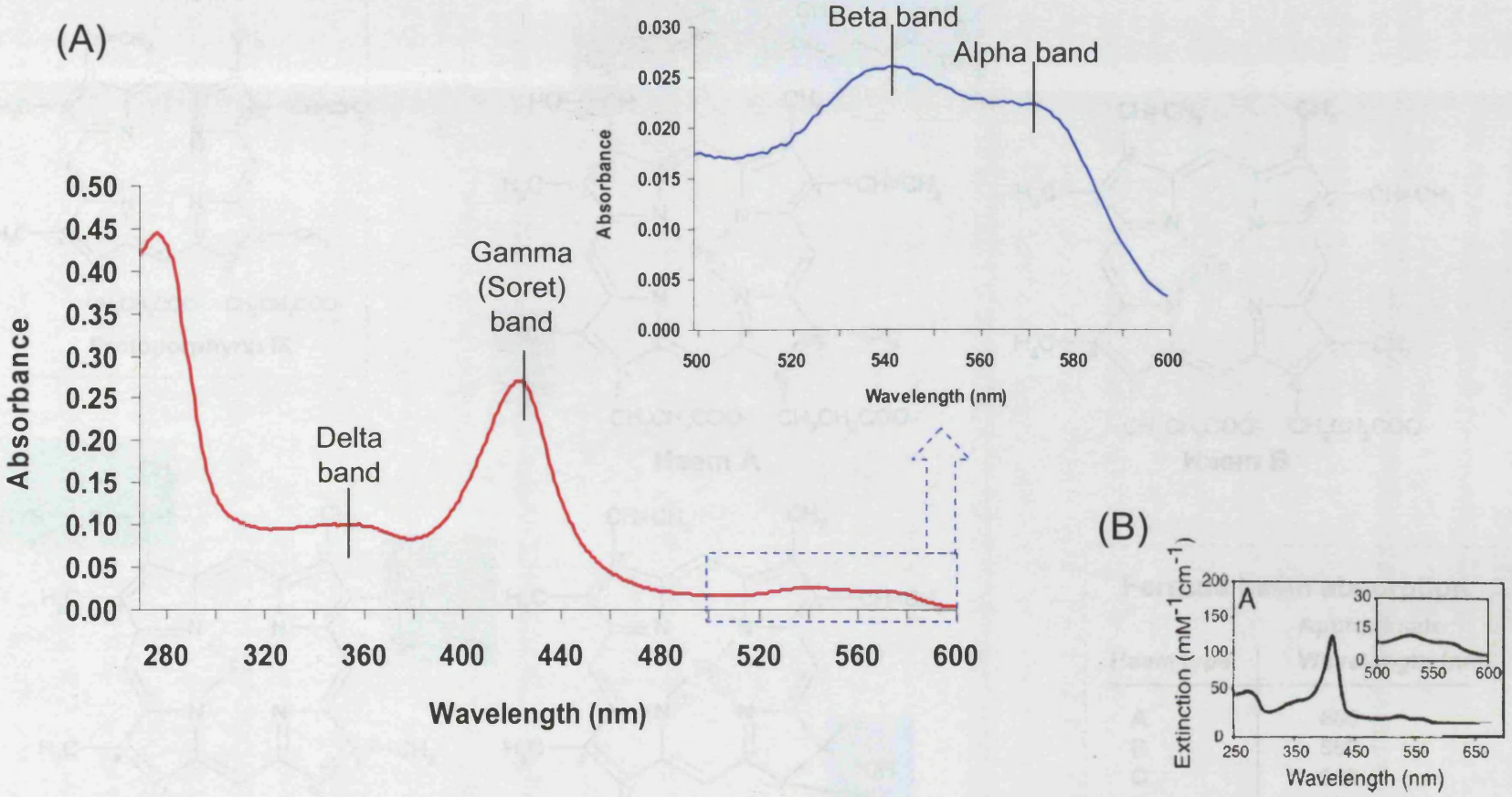


Figure 4-1:

(A) UV-Visible spectrum of $DosS_{1-379}$ recorded at room temperature, in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 % glycerol, 1 mM β-mercaptoethanol, in the range of 270-700 nm under aerobic conditions. The absorbance spectra show absorbance maxima at 424 nm, 541 nm and 571 nm, characteristic of a haem-binding protein. The insert is an enlargement of the range 500-600 nm.

(B) Absorption spectrum of DOS from *E. coli* (unrelated to *Mtb* DosS), a known haem-binding protein, adapted from Delgado-Nixon *et al.*, 2000.

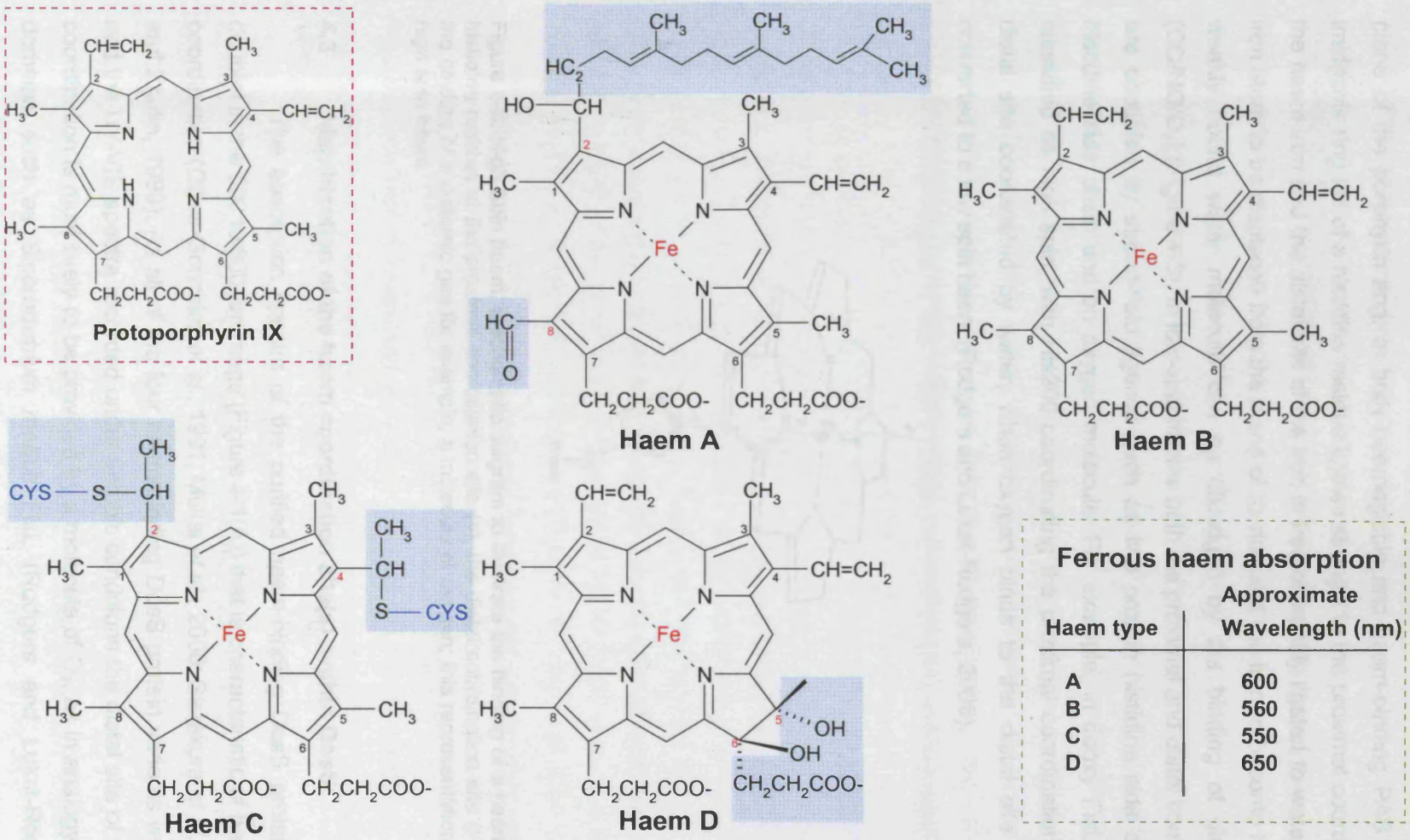


Figure 4-2: The structural differences between the different haem types. Type A found in cytochrome c oxidase; type B found in haemoglobin; type C found in cytochrome c and type D found in HPII catalase. All are derived from protoporphyrin IX (within the pink dashed box). The approximate characteristic absorption bands for haem in the reduced state, of each haem class, are listed within the green dashed box. Figure redrawn and adapted from (Neldon and Cox, 2005; Mahler and Cordes, 1969)

remaining fifth (proximal) and sixth (distal) coordination sites are perpendicular to the plane of the porphyrin ring. In both haemoglobin and haem-binding PAS domains, the imidazole ring N5 of a histidine residue ligates strongly to the proximal coordination site of the haem iron and the distal site of the iron is free or weakly ligated to water, causing the iron atom to be displaced from the plane of its ring and the iron to become high spin. The weakly bound water molecule can be dislodged by the binding of diatomic gases ($CO/NO/O_2$) (Figure 4-3). In low-spin haems both the proximal and distal coordination sites are occupied by strong-field ligands such as two protein histidine side chains or one histidine side chain and an oxygen molecule. For example, in deoxy FixL the haem is classified as high spin, with His200 coordinating the proximal coordination site and the distal site coordinated by water, when oxygen binds to the distal site the haem is converted to a low spin haem (Rodgers and Lukat-Rodgers, 2005).

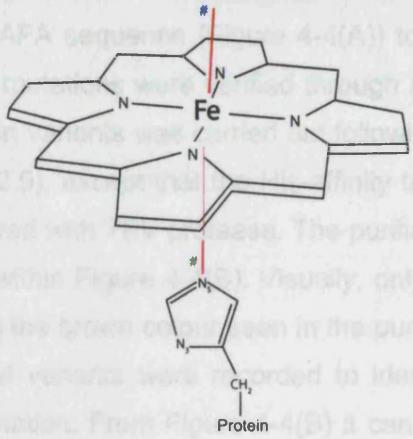


Figure 4-3: High Spin haem, a schematic diagram to illustrate the binding of a haem molecule to a histidine residue at the proximal coordination site (#), the distal coordination site (#) is left free for the binding of a diatomic gas for example, a molecule of oxygen; this representation is typical for a high spin haem

4.3 Identification of the haem coordination site(s) within DosS

The absorption spectra of the purified haem-binding DosS proteins showed a doublet in the 500 to 600 nm range (Figure 4-1(A)) that is characteristic of haem-iron hexa-coordination (Gilles-Gonzalez *et al.*, 1991; Mukai *et al.*, 2000; Sasakura *et al.*, 2002; Taylor and Zhulin, 1999). As all of the four haem-binding DosS protein variants were purified in and the UV-VIS spectra recorded under aerobic conditions the distal site of the haem iron coordination is most likely to be provided by a molecule of O_2, as in analogy to other PAS domains such as *Sinorhizobium meliloti* FixL (Rodgers and Lukat-Rodgers, 2005). However, there is a possibility that the distal coordination site may be provided by another

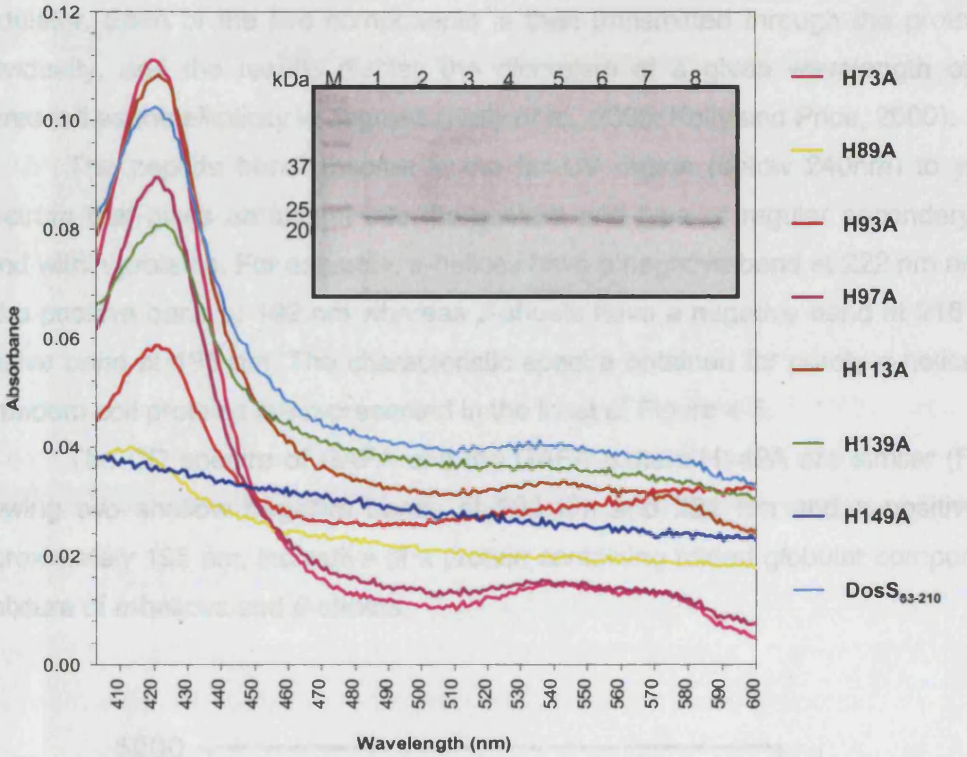
amino acid. For example, the absorption spectra recorded for *E. coli* Dos (EcDos; not related to mycobacterial DosS) indicated the presence of a six-coordinate haem (Figure 4-1(B)) but in this protein the sixth coordination site was provided by a methionine side chain (Gonzalez *et al.,* 2002; Sasakura *et al.,* 2002; Taylor and Zhulin, 1999). In most haemoproteins, such as haemoglobins, myoglobins, FixL and EcDos the proximal coordination site of the haem iron is a side-chain histidine amino acid.

In order to determine which histidine residue, (if any), of *MTB* DosS is involved in coordinating the haem-iron proximal site, site-specific mutagenesis was performed using a modified version of the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene), a PCR based protocol (section 2.4.8). GAFA, the shortest DosS variant, contains seven histidine residues and each of these individual residues were substituted with alanine. Alanine was chosen as it is a relatively inert amino acid, in both structural and functional terms. The single point mutations H73A, H89A, H93A, H97A, H113A, H139A and H149A were introduced into the GAFA sequence (Figure 4-4(A)) to create seven different constructs. All of the single point mutations were verified through sequencing (MWG) and purification of the individual protein variants was carried out following the same protocol used for wild-type protein (section 2.5), except that the His-affinity tag located at the N-terminus of the variants was not cleaved with TEV protease. The purified proteins were analysed by SDS-PAGE and is shown within Figure 4-4(B). Visually, only the purified H149A GAFA mutant was colourless, losing the brown colour seen in the purification of wild type protein. Visible spectra of the purified variants were recorded to identify and verify the involvement of H149 in haem coordination. From Figure 4-4(B) it can be seen that although the level of protein expression varied, thus the height of the absorption peaks differ, all mutants with the exception of H149A are still able to bind haem. This result suggests that H149 is either crucial in haem-binding or that the mutation causes an alteration in the protein structure, resulting in a loss of haem binding. The circular dichroism (CD) spectrum of the H149A single point mutation of GAFA was acquired and compared to the CD spectrum of wild-type GAFA to confirm whether the H149A protein is folded.

CD spectroscopy is a technique that provides confirmation of the folded status of a protein by measuring the difference in the absorption of plane polarised light. Plane polarised light is generated by two circularly polarised components of equal magnitude, one rotating anti-clockwise (left handed, L) and the other clockwise (right handed, R). When plane polarised light passes through a chiral chromophore, such as a protein, the L and R components are absorbed to different extents generating elliptically polarised radiation. Protein CD spectroscopy measures the difference in absorption between these

(A) DLEATLRAIVHSATSLVDARYGAMEVHDRQHRVLHFVYEGIDEETVRRIGHLPKGLGV
IGLLIEDPKPLRLDDVSAHPASIGFPPYHPPMRTFLGVPVRVRDESFGTLYLTDKTNG
QPFSDDDEVLVQALAAAAGIAVANARLYQQAK

(B)



(C)

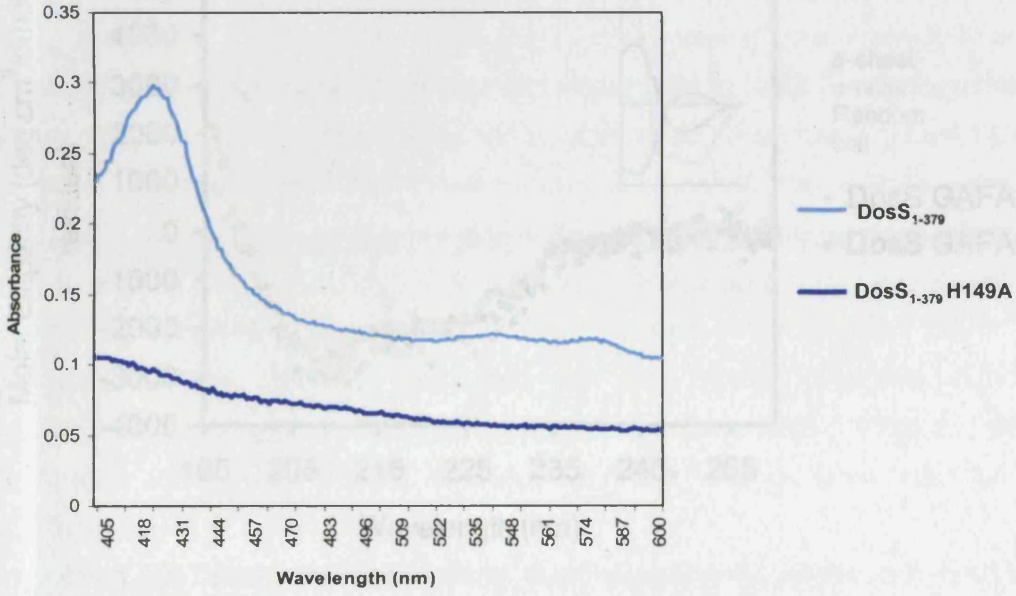


Figure 4-4: GAFA mutants

(A) Amino-acid sequence of GAFA, highlighted are the His to Ala mutations made.

(B) UV-VIS spectrum of the GAFA His to Ala mutants; inset:15 % SDS-PAGE analysis of 10 μg of the purified proteins. Lanes: M – molecular weight markers (kDa), 1 – GAFA, 2 – H73A, 3 – H89A, 4 – H93A, 5 – H97A, 6 – H113A, 7 – H139A, 8 – H149A. 10 μg of protein were loaded in each lane.

(C) UV-VIS spectrum of tandem GAFA GAFB construct $DosS_{1-379}$, and $DosS_{1-379}$ with His149 mutated to Ala. All spectra were recorded at room temperature, in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 % glycerol and 1 mM β-mercaptoethanol.

two components in the far-UV region (180-250 nm). In a spectropolarimeter the polarised light is split into the L and R circularly polarised components by passage though a modulator. Each of the two components is then transmitted through the protein sample individually, and the results display the dichroism at a given wavelength of radiation expressed as the ellipticity in degrees (Kelly *et al.*, 2005; Kelly and Price, 2000).

The peptide bond absorbs in the far UV region (below 240nm) to yield a CD spectrum that gives an insight into the content and type of regular secondary structure found within proteins. For example, α-helices have a negative band at 222 nm and 208 nm and a positive band at 192 nm whereas β-sheets have a negative band at 218 nm and a positive band at 196 nm. The characteristic spectra obtained for purely α-helical, β-sheet or random coil proteins are represented in the inset of Figure 4-5.

The CD spectra of GAFA and the GAFA mutant H149A are similar (Figure 4-5) showing two shallow negative bands at 208 nm and 222 nm and a positive band at approximately 195 nm, indicative of a protein containing folded globular components, with a mixture of α-helices and β-sheets.

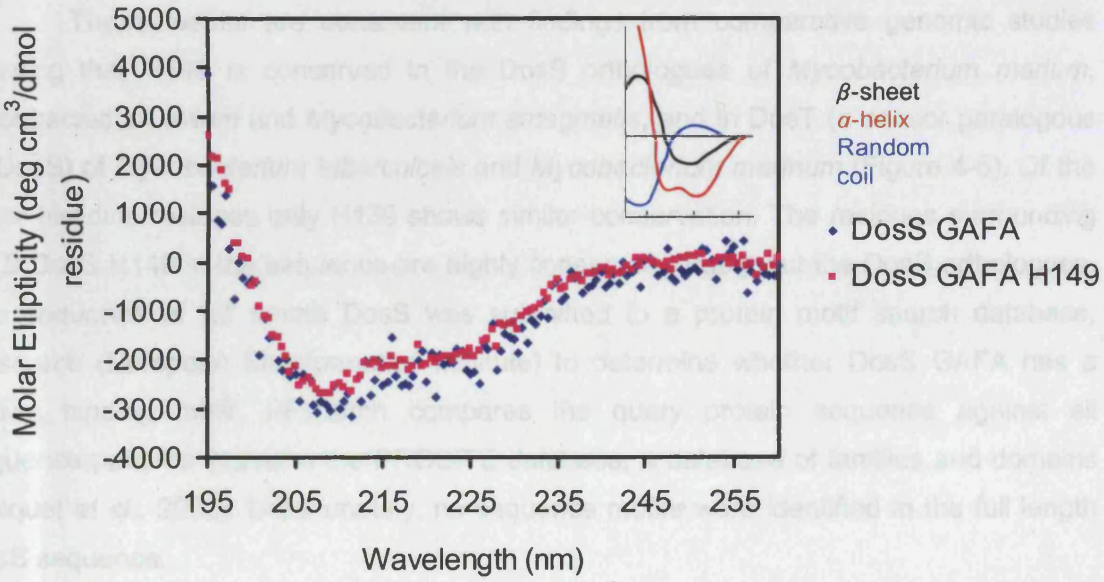


Figure 4-5: CD spectra of 14 μM GAFA H149A mutant and 7 μM GAFA recorded in 100 mM Tris pH 8.0 and 150 mM NaCl. The inset shows typical spectra for β-sheet, α-helical or random coil proteins

Therefore despite the variability in the levels of protein expression the specific point mutation H149A does not alter the overall structure of GAFA, as detected by CD spectroscopy, and thus H149 is the proximal coordination site of the haem iron.

Since only the single point mutation of H149A in GAFA alone is capable of abolishing the ability to bind haem it could be suggested that the distal coordination site of the haem co-factor is not occupied by another amino acid group from the protein. Otherwise it would be expected that the haem group would still be coordinated (although more weakly) to the protein through just the distal coordination site. Therefore, the distal coordination site is probably ligated to either water or a molecule of oxygen as the purification procedure was performed in aerobic conditions.

The single point mutation H149A was subsequently introduced into a construct containing both GAF domains $DosS_{1-379}$; after purification the protein variant $DosS_{1-379}$ H149A is not coloured and does not exhibit the haem spectral signature (Figure 4-1(C)), confirming that even within the context of the tandem GAF repeats of this sensory protein absence of the H149 side chain results in the loss of haem-binding. This observation reinforces the conclusion that haem binds only to the GAF domain proximal to the N terminus (GAFA). The presence of haem in *MTB* DosS is consistent with its role as a regulator of the hypoxia or latency response of *MTB* and suggests that signalling might be mediated through the presence of a high-spin haem cofactor. Interestingly, there are no histidine residues in the amino acid sequence of the DosS GAFB domain.

These results are consistent with findings from comparative genomic studies showing that H149 is conserved in the DosS orthologues of *Mycobacterium marium*, *Mycobacterium avium* and *Mycobacterium smegmatis*, and in DosT (a sensor paralogous to DosS) of *Mycobacterium tuberculosis* and *Mycobacterium marinum* (Figure 4-6). Of the other histidine residues only H139 shows similar conservation. The residues surrounding *MTB* DosS H149 in the sequence are highly conserved throughout the DosS orthologues. The sequence of full length DosS was submitted to a protein motif search database, PPsearch (European Bioinformatics Institute) to determine whether DosS GAFA has a known binding motif. PPsearch compares the query protein sequence against all sequence patterns stored in the PROSITE database, a database of families and domains (Falquet *et al.*, 2002). Unfortunately, no sequence motifs were identified in the full length DosS sequence.

DosT has been reported to undergo autophosphorylation at the conserved His392 residue followed by transfer of the phosphate group to a Asp54 on DosR (Roberts *et al.*, 2004; Saini *et al.*, 2004). The phosphorylation of Asp54 is essential for full activation of DosR under conditions of reduced oxygen tension (Roberts *et al.*, 2004). Domain analysis of *MTB* DosT using SMART (Schultz *et al.*, 1998) reveals the presence of two GAF domains, denoted A and B as depicted in Figure 4-7. Sequence comparisons

```
Mt_DosS   -------MTTGGLVDENDGAAMRPLRHTLSQLRLHELLVEVQDRVEQIVEGRDRLDGLVE 53
Mm_DosS   MTDEPHPQDAGPALGVPSGDGLRPLRLTLSQLRLRELLVEVQERVEQIVEGRDRLDGLVE 60
Msm_DosS  ------MGGVNGHHPRRDDDQRTPLRDTLSQLRLRELLSEVQDRVEQIVEGRDRLDGLVE 54
Mt_DosT   ---------VTHPDRANVNPGSPPLRETLSQLRLRELLLEVQDRIEQIVEGRDRLDGLID 51
Mm_DosT   ------------------------------------------MHAQLDELLAARDQMEQLLR 20
Ma_DosS   ---------------------------------------VTGNRFDELAAARDQTEKLLR 21
                                                    :.:::  .**: : *:

Mt_DosS   AMLVVTAGLDLEATLRAIVHSATSLVDARYGAMEVHDRQHRVLHFVYEGIDEETVRRIGH 113
Mm_DosS   AMLVVTSGLELNATLRTIVHSATNLVDARYCALEVHDRDKRVLQFVYEGIDEDTVARIGH 120
Msm_DosS  AMLVVTSGLELDVTLRTIVHTAIELVDARYGALGVRGEDHHLIEFIYEGITPEVREEIGP 114
Mt_DosT   AILAITSGLKLDATLRAIVHTAAELVDARYGALGVRGYDHRLVEFVYEGIDEETRHLIGS 111
Mm_DosT   VIVEIGSDLDLDATLRRIVLAARELTSAPYGALAVRDPEGTLIRFIHAGMDDDVARRIGH 80
Ma_DosS   VIAEIGAGLDLDATLHRIISAARELTSAPYGALAVRDWHANLISFVHEGMDAETVRRIGH 81
          .:  : :.*.*:.**: *: :* .*..* * *: *:. .  :: *:: *:  :.  **

Mt_DosS   LPKGLGVIGLLIEDPKPLRLDDVSAHPASIGFPPYHPPMRTFLGVPVRVRDESFGTLYLT 173
Mm_DosS   LPEGLGVIGLLIDEPKPLRLEDISQHPASIGFPSHHPPMRTFLGVPVRVRDVSLGTLYLT 180
Msm_DosS  LPEGRGVLGVLIDDPKPIRLDDIRLHPASVGFPPNHPPMRTFLGVPVRTRDKVFGNLYLT 174
Mt_DosT   LPEGRGVLGALIEEPKPIRLDDISRHPASVGFPLHHPPMRTFLGVPVRIRDEVFGNLYLT 171
Mm_DosT   LPVGKGVLGVSLVETPALRLDDLTMHPAACGFPEHHPPMRAFLAVPITIRGTVFGSLYLT 140
Ma_DosS   HPVGKGLLSLSLLDTPALRMDDLTAHPAAAGFPEHHPAMRAFLAVPITIRGAVFGNLYLT 141
           * * *::.  : :. .:*::*: ***:  ***  **.**:**.**:  *.  :*.****

Mt_DosS   DKTNGQPFSDDDEVLVQALAAAAGIAVANARLYQQAKARQSWIEATRDIATELLSGTEP- 232
Mm_DosS   DKTNGQPFSDDDELLVQALAAAAGIAIANARLYEQAKARQAWIGATRDIATELLSGTEP- 239
Msm_DosS  EKTNGQPFSEDDEVLVQALAAAAGIAVDNARLYEESQARQAWIAATRDIGTQLLSGTDP- 233
Mt_DosT   EKADGQPFSDDDEVLVQALAAAAGIAVDNARLFEESRTREAWIEATRDIGTQMLAGADP- 230
Mm_DosT   HVDPGRKFSESDEFAARALAFAAGVAIDNAQVFERERTAVKWMEASREITTALMSSTESR 200
Ma_DosS   HVDEARVFSDADEMAARALAFAAAVAIDNAQVFERERMSVKWIEASREITTALLSRTEPH 201
          .  .: **: **. .:*** **.:*: **:::::. :      *: *:*:* * ::: ::.
```

Figure 4-6: Sequence comparison of GAFA domains in DosS of *Mycobacterium tuberculosis* (*Mt*) and its orthologues in *Mycobacterium marinum* (*Mm*), *Mycobacterium avium* (*Ma*) and *Mycobacterium smegmatis* (*Msm*), and in DosT of *Mt* and *Mm*. Residues in blue correspond to the putative domain GAFA used in this study. All histidine residues are coloured red while the conserved histidine residue H149 in *Mtb* is shown using bold font. Stars annotate strictly conserved residues in all six sequences while dots represent level of similarity.

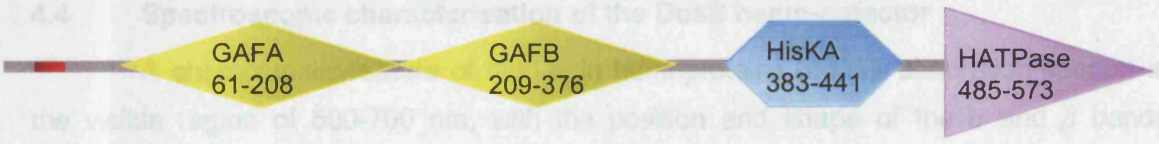


Figure 4-7: Domain organisation of DosT as predicted by SMART (Schultz *et al.*, 1998)..

between DosT and DosS for the individual GAFA and GAFB domains using the program EMBL-EBI ClustalW (Thompson *et al.*, 1994) show that they have a 72 % and 52 % sequence identity, respectively. Interestingly, there is a 58 % sequence identity between the full length amino acid sequences of DosT and DosS. The *in vivo* roles of DosS and DosT in *MTB* appear to overlap as both are involved in the *MTB* genetic response to hypoxia. It is therefore possible that DosT also functions by the presence of a haem cofactor. To investigate this further the GAFA and GAFB domains and full length DosT DNA constructs have been cloned, although expression tests have not yet been initiated.

4.3.1 Haem-incorporation

Although the haem-binding DosS proteins are able to incorporate endogenous haem produced by *E. coli*, the rate of protein synthesis is often higher than the rate of haem biosynthesis (Varnado and Goodwin, 2004) leading to a proportion of the protein lacking the haem prosthetic group required for activity. As the DosS proteins were being subjected to crystallisation trials (section 2.7.5) it was important to produce homogeneous protein samples. The presence of mixed populations of protein is known to hinder the formation of crystals, as molecules can assembly poorly in a crystal lattice producing low quality diffraction data (Pusey *et al.*, 2005).

In order to promote maximum incorporation of the haem cofactor into the proteins, the expression procedure was altered to include the addition of δ-aminolevulinic acid, the product of the first step in the haem-biosynthetic pathway, to enhance the endogenous levels of haem produced from *E. coli* for uptake in protein expression. Also haemin (iron protophyrin chloride) was added to the cell lysates to ensure optimal incorporartion of haem cofactor (section 2.6.4). This procedure resulted in an increase in absorbance at 424 nm relative to that at 280 nm for all of the haem containing DosS proteins (GAFA, $DosS_{1-379}$ and full length DosS). The 424/280 ratio of $DosS_{1-379}$ increased from 0.37 to 0.68 (Figure 4-8), thus approximately doubling the haem incorporation, with similar increments were observed for the other DosS haem-binding protein constructs.

4.4 Spectroscopic characterisation of the DosS haem-cofactor

A characteristic feature of high-spin hemoproteins is their absorption spectra in the visible region of 500-700 nm, with the position and shape of the α and β bands dependent upon haem type (A, B, C and D), redox state and ligand present. UV-VIS spectroscopy was utilised to characterise the redox state of the DosS haem co-factor and

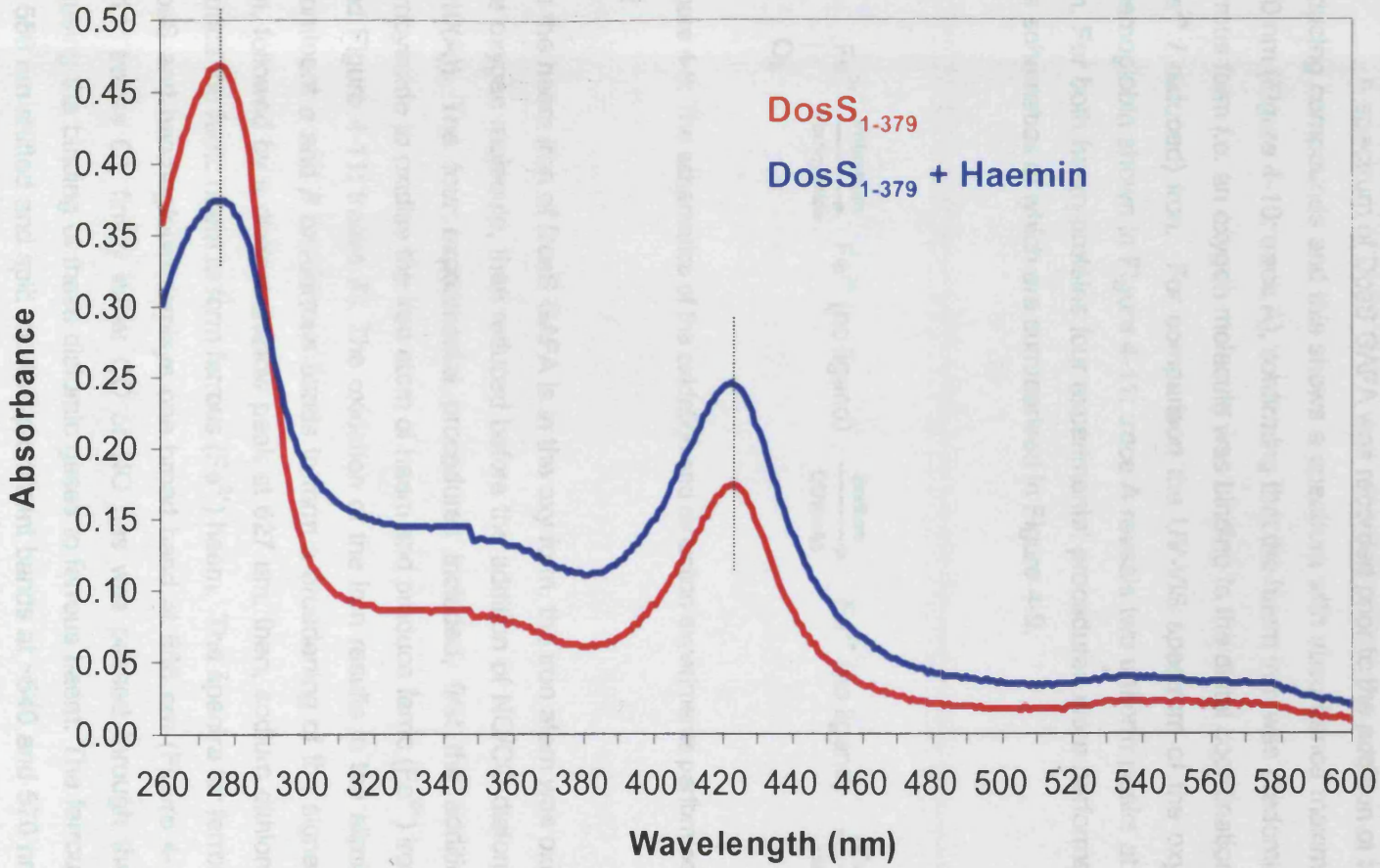


Figure 4-8: Recorded UV-visible absorbance spectra of $DosS_{1-379}$ purified with and without the addition of haemin, at room temperature, in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 % glycerol, 1 mM β-mercaptoethanol in the range of 260-600 nm. The addition of haemin causes an increase in the radio of 424: 280 nm (37% to 68%).

its ability to bind diatomic gases to the distal site of coordination. For comparative purposes the UV- VIS spectra of haemoglobin, which contains a classic high spin b-type haem that is able to bind diatomic gases, were acquired in parallel to that of DosS.

A spectrum of DosS GAFA was recorded prior to the addition of haem oxidising or reducing compounds and this shows a spectrum with absorbance maxima at 542 nm and 570 nm (Figure 4-10: trace A), confirming that the haem iron was predominantly in the oxy-ferrous form *i.e.* an oxygen molecule was binding to the distal coordination site of a ferrous (Fe^{2+} / reduced) iron. For comparison the UV-VIS spectrum of the oxy-ferrous state of haemoglobin shown in Figure 4-11: trace A reveals two uniform peaks at 542 nm and 577 nm. For both haem-proteins four experimental procedures were performed (section 2.7.7) the schematics of which are summarised in Figure 4-9.

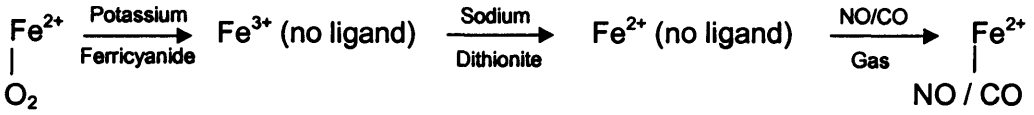


Figure 4-9: The schematics of the oxidation and reduction experiments performed on DosS GAFA.

As the haem iron of DosS GAFA is in the oxy form, the iron atom was oxidised to displace the oxygen molecule, then reduced before the addition of NO/CO diatomic gases (Figure 4-10(A)). The four experimental procedures included, first the addition of potassium ferricyanide to oxidise the iron atom of haem and produce ferric (Fe^{3+}) iron (Figure 4-10(B) and Figure 4-11: traces B). The oxidation of the iron results in the elimination of the two prominent α and β oxy-ferrous bands to form a broadening of the signal from 550 – 600 nm, followed by a distinct shallow peak at 627 nm; then, sodium dithionite was added to reduce the ferric haem to form ferrous (Fe^{2+}) haem. The spectra for ferrous (Fe^{2+}) haem in DosS and haemoglobin displays one broad band at 556 nm (Figure 4-10(B) and Figure 4-11: traces C); finally either CO or NO gas was passed through the protein sample causing the binding of these diatomic gases to ferrous haem. The ferrous band produced at 556 nm shifted and split into two prominent bands at ~540 and 570 nm (Figure 4-10(B) and Figure 4-11: traces D and E). Table 4-1 summaries the UV-VIS spectroscopic data obtained for DosS GAFA (Figure 4-10(B)) and haemoglobin (Figure 4-11). Interestingly, after the addition of NO the maxima in the UV-VIS spectra of DosS and haemoglobin are

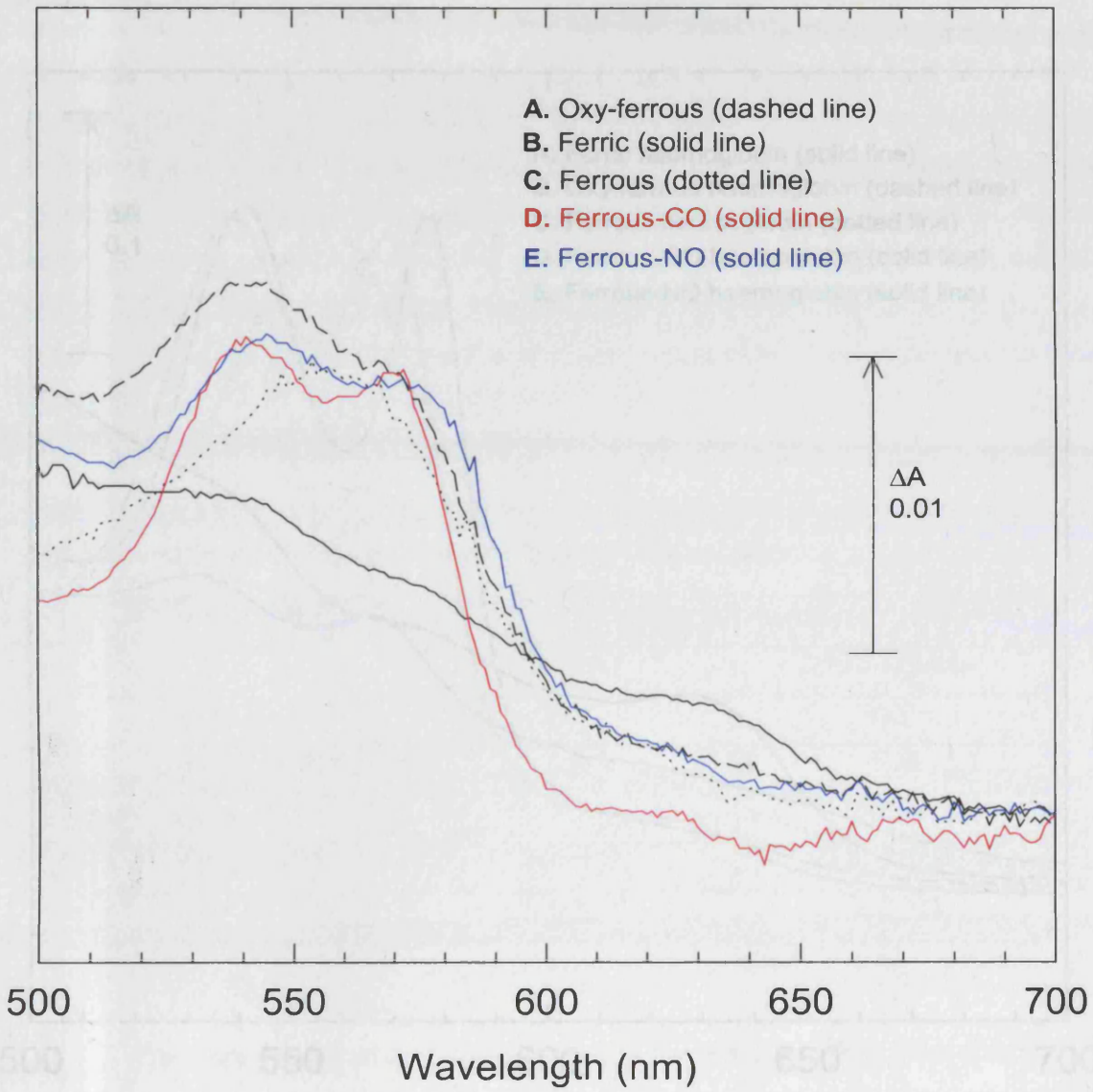


Figure 4-10: UV-VIS spectra of DosS GAFA
(A) Absorbance spectrum of 5 μM DosS GAFA recorded in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM β-mercaptoethanol, 5 % glycerol
(B) plus the addition of 1.5 μl 500 mM potassium ferricyanide
(C) or solid sodium dithionite
(D) CO gas
(E) NO gas.

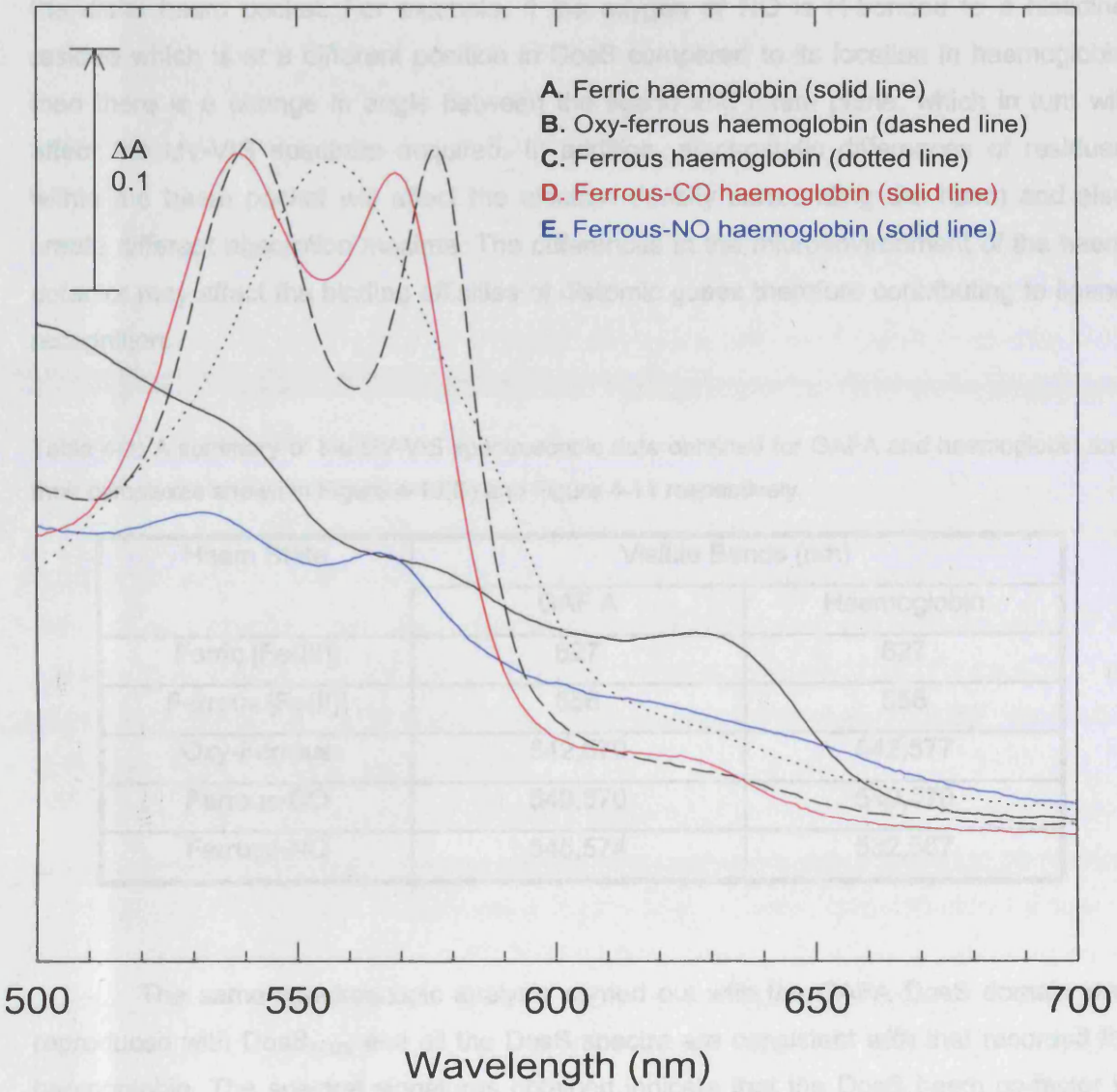


Figure 4-11: Absorption spectra of 5 mM haemoglobin recorded in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM β-mercaptoethanol, 5 % glycerol. Spectrum recorded after the addition of:

(A) 250 mM potassium ferricyanide
(B) 5 mM sodium dithionite
(C) 20 mM sodium dithionite
(D) CO gas
(E) NO gas.

different. In DosS GAFA the binding of NO shifts the spectrum to the left of the spectra obtained for the oxy-ferrous and ferrous-CO haem states, whereas in haemoglobin the shift is to the right. The difference in the wavelength maxima of the NO binding spectra between these two proteins probably arises from the geometry of the bound ligand within the distal haem pocket. For example, if the oxygen of NO is H-bonded to a histidine residue which is at a different position in DosS compared to its location in haemoglobin then there is a change in angle between the ligand and haem plane, which in turn will affect the UV-VIS spectrum acquired. In addition, electrostatic differences of residues within the haem pocket will affect the electron density surrounding the haem and also create different absorption maxima. The differences in the microenvironment of the haem cofactor may affect the binding affinities of diatomic gases therefore contributing to ligand recognition.

Table 4-1: A summary of the UV-VIS spectroscopic data obtained for GAFA and haemoglobin and their complexes shown in Figure 4-10(B) and Figure 4-11 respectively.

Haem State	Visible Bands (nm)	
	GAF A	Haemoglobin
Ferric [Fe(III)]	627	627
Ferrous [Fe(II)]	556	556
Oxy-Ferrous	542,570	542,577
Ferrous-CO	540,570	540,570
Ferrous-NO	545,574	532,567

The same spectroscopic analysis carried out with the GAFA DosS domain was reproduced with $DosS_{1-379}$ and all the DosS spectra are consistent with that recorded for haemoglobin. The spectral signatures obtained indicate that the DosS haem co-factor is able to bind classic haem ligands to its sixth co-ordination site indicative of a high-spin haem. The prominent broad spectral band at 556 nm in the ferrous (reduced) state suggests the presence of a b type haem, (the a, c and d ferrous haems typically have spectral bands at 600, 550 and 650 nm, respectively).

The two component system DosRS has been implicated in the adaptation of *MTB* to unfavourable hypoxic and NO conditions. The UV-VIS spectroscopic data for DosS presented herein clearly establishes that its haem has the ability to produce stable complexes with O_2, CO and NO (Figure 4-10(B): traces A, D, E) consistent with its

potential role as a gas sensor *in vivo*. Prior to these studies the actual molecular nature of the DosS stimuli was not known, although Park *et al.* commented that DosR was necessary for the induction of a genetic response to reduced oxygen tension and intuitively suggested that a haem protein would play a role in monitoring the available oxygen (Park *et al.*, 2003).

4.5 Structural investigation of GAFA DosS

The structural database SMART predicts that there are 5286 GAF domains within 4694 proteins, belonging to all kingdoms of life (Schultz *et al.*, 1998). It is thought that these domains bind an array of small molecules to regulate signalling events (Zoraghi *et al.*, 2004), however direct evidence of ligand binding has only been demonstrated for a few GAF domains. The most commonly reported ligands are cyclic nucleotides as in cyclic nucleotide phosphodiesterases (PDEs). Other ligands shown to bind to GAF domains are formate, 2-oxoglutarate, bilin and pyruvate (Zoraghi *et al.*, 2004). The GAFA domain of DosS is the first GAF domain reported to contain a haem cofactor. Despite the large number of predicted GAF domains only five structures have been published, all solved by X-ray crystallography. These GAF domain structures contain a 6 stranded anti-parallel β-sheet with strand order 3-2-1-6-5-4 surrounded by at least 4 α-helices (Figure 4-12).The structural database SCOP (Murzin *et al.*, 1995) classifies the core of a GAF domain as consisting of an α-α-β-β-β-α-β-β-β-α fold.

In order to investigate the structural similarity of the DosS GAF domains to other protein domains, a robust sequence search was carried out using the SAMt99 method (Karplus and Hu, 2001), which utilises a hidden Markov model-based protocol to detect remote homologues of the full length DosS sensory protein. From the 11149 homologies identified, 24 have a known 3D structure. None of the identified homologues produced significant global alignment (>20% sequence similarity), only local alignments that cover either the N- or C- terminal regions and furthermore, no homologue was found that covered both putative GAF domains. Among the identified sequence homologues, no structural homologue was identified for DosS GAFB, and only one, structural homologue found for DosS GAFA. The GAF B domain from cyclic nucleotide phosphodiesterase 2A (PDE2A) (pdb code 1MC0 (Martinez *et al.*, 2002)) shares 25 % sequence identity to DosS GAFA and, therefore, is most likely to share a similar secondary structure topology. Figure 4-13 shows the sequence alignment between the first GAF domain of *M. tuberculosis*

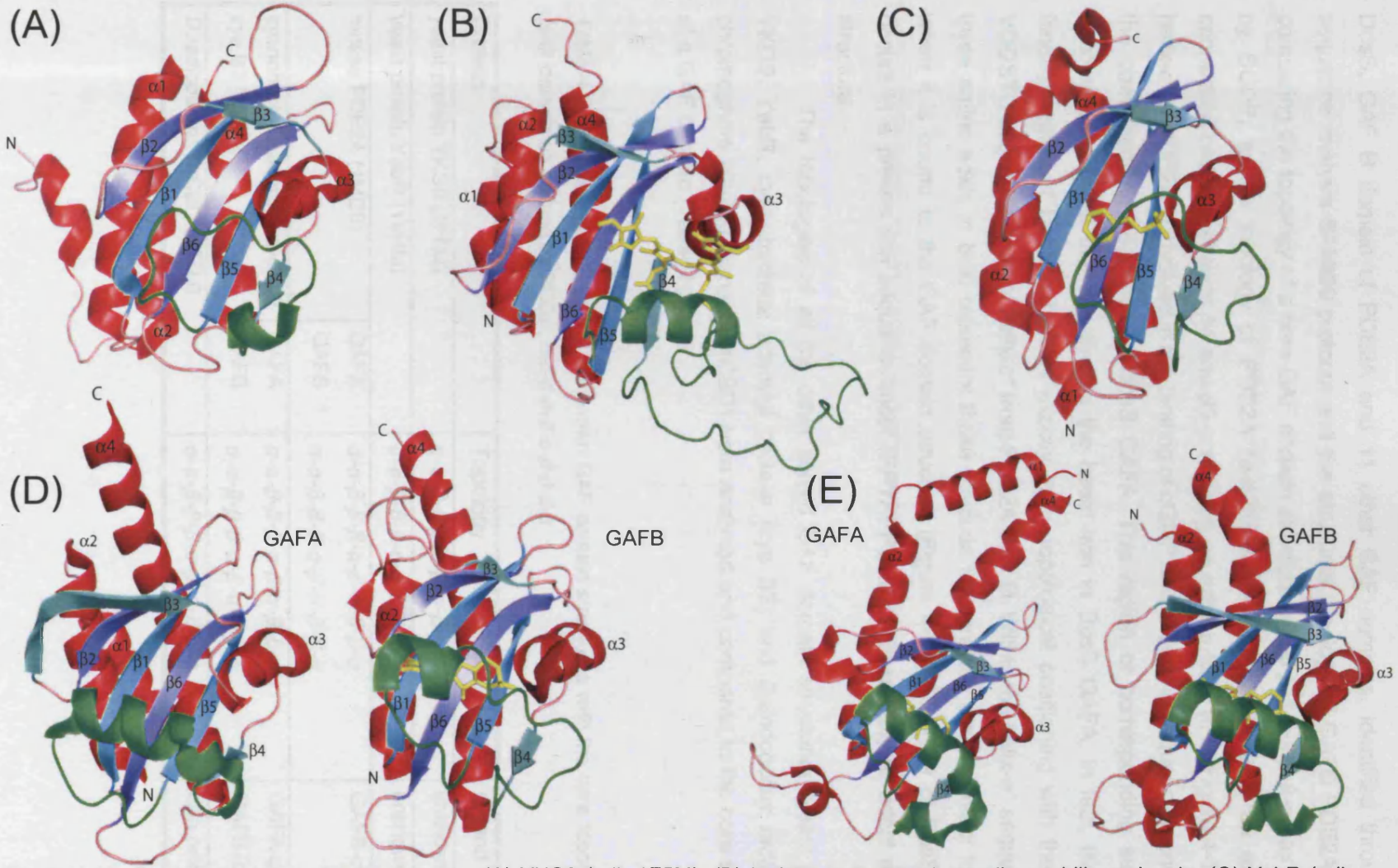


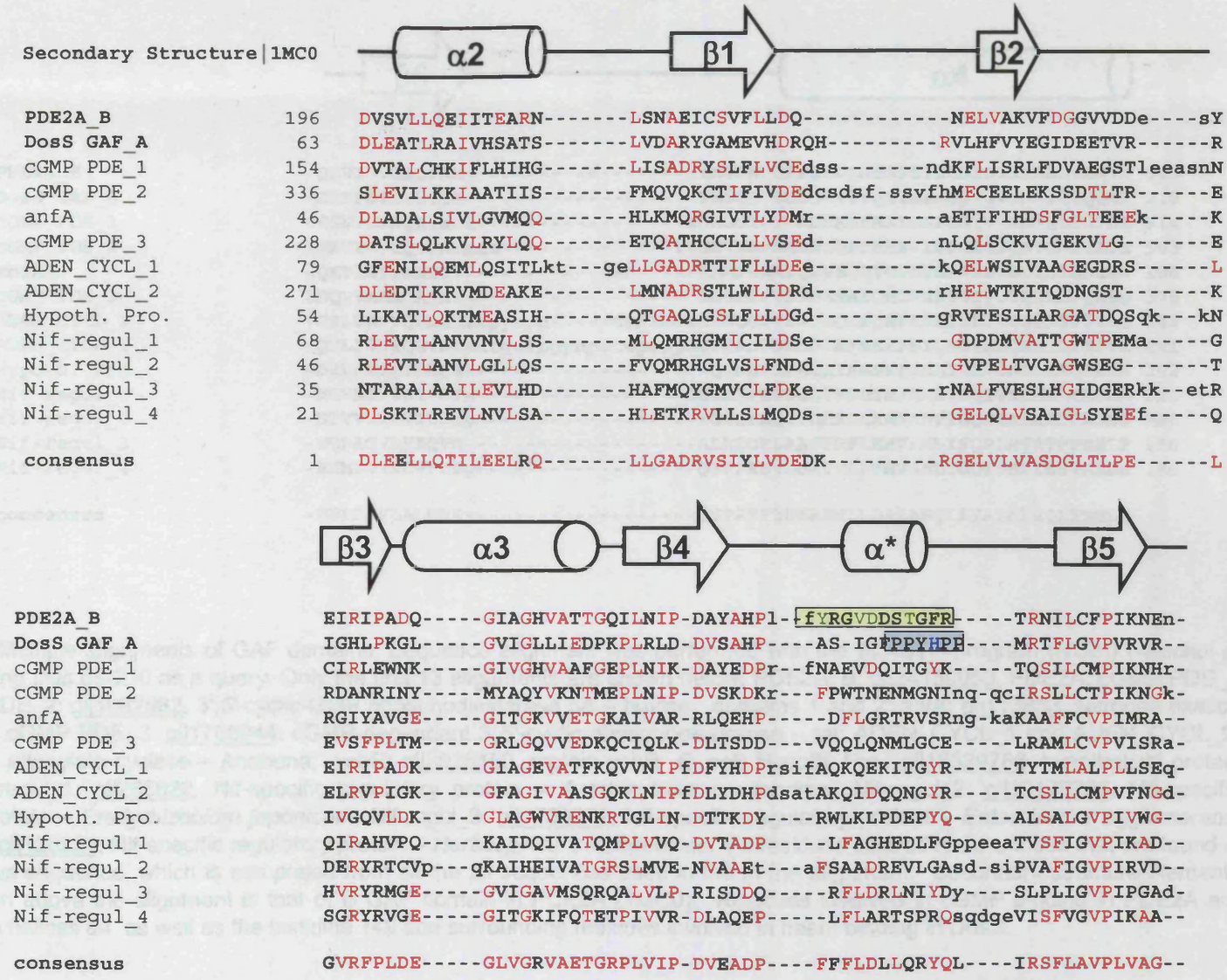
Figure 4-12: Crystal structures of GAF domains: (A) YKG9 (pdb 1F5M), (B) DrCBC (pdb 1ZTU) in yellow a bilin molecule, (C) YebR (pdb 1VHM), in yellow a Mes molecule, (D) PDE2 (pdb 1MC0), GAFB solved complexed to cGMP (yellow), (E) CynB2 (pdb 1YKD), GAFA and GAFB complexed to cAMP (yellow). Helix-α^* is coloured green; the structures are not drawn to scale.

DosS, GAF B domain of PDE2A and 11 other GAF domains, identified through the sequence analysis SAMt99 protocol and the structural topology of GAFB PDE2A. When comparing the topology of a core GAF domain structure (*α-α-β-β-β-α-β-β-β-α*), as defined by SCOP, to the topology of PDE2A (*α-α-β-β-β-α-β-α-β-β-α*), there is an extended connection between strands *β*4 and *β*5 containing an extra *α*-helix (*α**) (Figure 4-13). The helix-*α** of PDE2A is involved in the binding of cGMP and shows no sequence homology to the corresponding sequence of DosS GAFA. This region of corresponding sequence encompasses H149 that coordinates the haem iron in DosS GAFA. In fact, the haem binding region of DosS correlates exactly in its topological positioning with the cGMP **VD**DSTG sequence motif of helix-*α** from PDE2A GAF B. Within the helix-*α** sequence the three amino acids in bold represent those residues that provide direct contact to cGMP when it is bound to the GAF domain structure (Figure 4-14(A)). H149 of DosS GAFA occurs in a proline rich sequence motif (PPYHPP), which is unlikely to adopt a helical structure.

The topologies of all the other solved GAF domain structures (yeast proteins YKG9, YebR, cyanobacteria adenylyl cyclase (cya B2) and *Deinococcus radiodurans* chromophore binding domain (DrCBD) were analysed and compared to the core topology of a GAF domain (Table 4-2).

Table 4-2: Topology comparison of the known GAF domain structures with the core topology of a GAF domain as defined by SCOP, *α-α-β-β-β-α-β-β-β-α*.

Protein		Topology	Ligand
yeast protein YKG9 (1F5M)		*α-α-β-β-β-α-β-α-β-β-α*	unknown
yeast protein YebR (1VHM)		*α-α-β-β-β-α-β-β-β-α*	unknown
mouse PDE2A (1MC0)	GAFA GAFB	*α-α-β-β-β-α-β-α-β-β-α* *α-α-β-β-β-α-β-α-β-β-α*	GAFB-cGMP,
cyanobacteria adenylyl cyclase cya B2 (1YKD)	GAFA GAFB	*α-α-β-β-β-α-β-α-β-β-α* *α-α-β-β-β-α-β-α-β-β-α*	GAFA cAMP, GAFB- cAMP
D.radiodurans DrCBD (1ZTU)		*α-α-β-β-β-α-β-α-β-β-α*	bilin molecule


Secondary Structure|1MC0
α2
β1
β2
β3
α3
β4
α*
β5
PDE2A_B 196 DVSVLLQEIITEARN-------LSNAEICSVFLLDQ-----------NELVAKVFDGGVVDDe----sY
DosS GAF_A 63 DLEATLRAIVHSATS-------LVDARYGAMEVHDRQH---------RVLHFVYEGIDEETVR------R
cGMP PDE_1 154 DVTALCHKIFLHIHG-------LISADRYSLFLVCEdss--------ndKFLISRLFDVAEGSTleeasnN
cGMP PDE_2 336 SLEVILKKIAATIIS-------FMQVQKCTIFIVDEdcsdsf-ssvfhMECEELEKSSDTLTR------E
anfA 46 DLADALSIVLGVMQQ-------HLKMQRGIVTLYDMr----------aETIFIHDSFGLTEEEk-----K
cGMP PDE_3 228 DATSLQLKVLRYLQQ-------ETQATHCCLLLVSEd----------nLQLSCKVIGEKVLG-------E
ADEN_CYCL_1 79 GFENILQEMLQSITLkt---geLLGADRTTIFLLDEe----------kQELWSIVAAGEGDRS------L
ADEN_CYCL_2 271 DLEDTLKRVMDEAKE-------LMNADRSTLWLIDRd----------rHELWTKITQDNGST-------K
Hypoth. Pro. 54 LIKATLQKTMEASIH-------QTGAQLGSLFLLDGd----------gRVTESILARGATDQSqk---kN
Nif-regul_1 68 RLEVTLANVVNVLSS-------MLQMRHGMICILDSe----------GDPDMVATTGWTPEMa-----G
Nif-regul_2 46 RLEVTLANVLGLLQS-------FVQMRHGLVSLFNDd----------GVPELTVGAGWSEG-------T
Nif-regul_3 35 NTARALAAILEVLHD-------HAFMQYGMVCLFDKe----------rNALFVESLHGIDGERkk--etR
Nif-regul_4 21 DLSKTLREVLNVLSA-------HLETKRVLLSLMQDs----------GELQLVSAIGLSYEEf-----Q
consensus 1 DLEELLQTILEELRQ-------LLGADRVSIYLVDEDK---------RGELVLVASDGLTLPE------L
PDE2A_B EIRIPADQ-----GIAGHVATTGQILNIP-DAYAHPl-fYRGVDDSTGFR-----TRNILCFPIKNEn-
DosS GAF_A IGHLPKGL-----GVIGLLIEDPKPLRLD-DVSAHP----AS-IGIPPYHPE----MRTFLGVPVRVR--
cGMP PDE_1 CIRLEWNK-----GIVGHVAAFGEPLNIK-DAYEDPr--fNAEVDQITGYK-----TQSILCMPIKNHr-
cGMP PDE_2 RDANRINY-----MYAQYVKNTMEPLNIP-DVSKDKr---FPWTNENMGNInq-qcIRSLLCTPIKNGk-
anfA RGIYAVGE-----GITGKVVETGKAIVAR-RLQEHP-----DFLGRTRVSRng-kaKAAFFCVPIMRA--
cGMP PDE_3 EVSFPLTM-----GRLGQVVEDKQCIQLK-DLTSDD----VQQLQNMLGCE-----LRAMLCVPVISRa-
ADEN_CYCL_1 EIRIPADK-----GIAGEVATFKQVVNIPfDFYHDPrsifAQKQEKITGYR-----TYTMLALPLLSEq-
ADEN_CYCL_2 ELRVPIGK-----GFAGIVAASGQKLNIPfDLYDHPdsatAKQIDQQNGYR-----TCSLLCMPVFNGd-
Hypoth. Pro. IVGQVLDK-----GLAGWVRENKRTGLIN-DTTKDY----RWLKLPDEPYQ-----ALSALGVPIVWG--
Nif-regul_1 QIRAHVPQ-----KAIDQIVATQMPLVVQ-DVTADP-----LFAGHEDLFGppeeaTVSFIGVPIKAD--
Nif-regul_2 DERYRTCVp---qKAIHEIVATGRSLMVE-NVAAEt---aFSAADREVLGAsd-siPVAFIGVPIRVD--
Nif-regul_3 HVRYRMGE-----GVIGAVMSQRQALVLP-RISDDQ-----RFLDRLNIYDy----SLPLIGVPIPGAd-
Nif-regul_4 SGRYRVGE-----GITGKIFQTETPIVVR-DLAQEP-----LFLARTSPRQsqdgeVISFVGVPIKAA--
consensus GVRFPLDE-----GLVGRVAETGRPLVIP-DVEADP----FFFLDLLQRYQL----IRSFLAVPLVAG--


```
PDE2A_B          -QEVIGVAELVNK------------------INGPWFSKFDEDLATAFSIYCGISIAHSLLYKKVN 345
DosS GAF_A       -DESFGTLYLTDK------------------TNGQPFSDDDEvlvqalaaaagiavanarlYQQAK 210
cGMP PDE_1       -EEVVGVAQAINKk-----------------sGNGGTFTEKDEKDFAAYLAFCGIVLHNAQLYETSL 314
cGMP PDE_2       kNKVIGVCQLVNKmee--------------ttGKVKAFNRNDEQFLEAFVIFCGLGIQNTQMYEAVE 503
anfA             -QKVLGTIAAERV------------------YMNPRLLKQDVELLTMIATMIAPLVELYLIENIER 196
cGMP PDE_3       tDQVVALACAFNK------------------LGGDFFTDEDERAIQHCFHYTGTVLTSTLAFQKEQ 375
ADEN_CYCL_1      -GRLVAVVQLLNKlkpyspp-----dallaerIDNQGFTSADEQLFQEFAPSIRLILESSRSFYIAT 249
ADEN_CYCL_2      -QELIGVTQLVNKkktgefppynpetwpiapeCFQASFDRNDEEFMEAFNIQAGVALQNAQLFATVK 441
Hypoth. Pro.     -DELLGILTLMHS------------------QVNHFTPACATAMEKTAELIALVLNNARIQTKHK 202
Nif-regul_1      -HHVMGTLSIDRIw-----------------dGTARFRFDEDVRFLTMVANLVGQTVRLHKLVASDR 220
Nif-regul_2      -STVVGTLTIDRIp-----------------EGSSSLLEYDARLLAMVANVIGQTIKLHRLFAGDR 198
Nif-regul_3      -NQPAGVLVAQPM------------------ALHEDRLAASTRFLEMVANLISQPLRSATPPESLP 186
Nif-regul_4      -REMLGVLCVFRDg-----------------QSPSRSVDHEVRLLTMVANLIGQTVRLYRSVAAER 180

consensus        -GELLGVLALHRK------------------DSPRPFTEEEEELLQALANQLAIALALAQLYEELR
```

Figure 4-13: Multiple alignments of GAF domains. Sequence alignment was performed with the BLASTP program (NCBI) (Altschul et al., 1997) using Dos 63-210 as a query. Only the first 13 alignments are shown below: PDE2A_B: gi|24159053, PDE2A; cGMP PDE_1 and cGMP PDE_2: gi|1082982, 3',5'-cyclic-GMP phosphodiesterase 5A – bovine, domains 1 and 2; anfA: gi|113833, Nitrogen fixation protein anfA; cGMP PDE_3: gi|1705944: cGMP-dependent 3',5'-cyclic phosphodiesterase – rat; ADEN_CYCL_1 and ADEN_CYCL_2: gi|17229396, adenylate cyclase – Anabena; yebR: gi|3915450, protein yebR- *E. coli*; Hypoth. Pro.: gi|16329786, hypothetical protein slr1385; Nif-regul_1: gi|266622, Nif-specific regulatory protein – *Azorhizobium caulinodans*; Nif-regul_2: gi|30179736, Nif-specific regulatory protein – *Bradyrhizobium japonicum*; Nif-regul_3: gi|1709272, Nif-specific regulatory protein – *Enterobacter agglomerans*; Nif-regul_4: gi|128221, Nif-specific regulatory protein – *Herbaspirillum seropedicae*. All residues colored red are those that are found in the consensus sequence, which is comprised from all the all sequences used in the in the alignment. Secondary structure elements representation above the alignment is that of B GAF domain in PDE2A (1MC0). Residues involved in cGMP binding in PDE2A are highlighted in helices *α*4, as well as the histidine 149 and surrounding residues involved in haem binding in DosS.

From the comparison of topologies, it can be seen that all of the GAF domain structures, with the exception of YebR, contain the extra helix (α^*) between strands β-4 and β-5 (Figure 4-12), consistent with the structure of PDE2A. In fact, in both the structures of cyn2a and DrCBD, which were solved complexed to a ligand, the additional helix is also involved in binding (Figure 4-14).

Within the fold of these GAF domains the 6-stranded anti-parallel β-sheet forms the roof of the binding pocket and helices α3 and α^* form the crevice of the binding cavity (Figure 4-14). All of the ligands are located on the same face of the central β-sheet with the N- and C- terminal helices positioned on the opposite side to the binding site. The location of helix α3, on one side of the binding region is consistent in all GAF domain structures. The helix-α^* is variable in length but is found to pack against the ligand, on the opposite side of the sheet to helix α3. Helix-α^* is connected to strand β5 via a loop that also forms contacts with the ligands. This loop together with helix-α^* is known as the variable region or 'irregular layer' of a GAF domain (Ho *et al.*, 2000). Therefore, it seems probable that this loop region and helix-α^* form the contacts that confer specificity needed for ligand binding. The variability in the length of the loop region and helix-α^* may provide flexibility to accommodate a variety of sized ligands. Figure 4-14 illustrates the interactions of the GAF ligands with helix-α3 and the variable region for cya B2, DrCBD and PDE2A. The crystal structure of YebR does not contain this extra helix-α^* but instead a longer loop region connecting strands β4 to β5. The structure of YKG9 which was solved without a bound ligand also has a long loop region in between these β-strands (Figure 4-12), and it may be possible that this loop region has a propensity to form a helical conformation upon the binding of a ligand.

The secondary structure topology of DosS GAFA, predicted using PSIPRED (McGuffin *et al.*, 2000), consists of an α-β-β-α-α-β-α-β-β-α fold (Figure 4-15). PSIPRED uses sequence profile alignments in order to determine putative β-sheets and α-helices, with an accuracy of 77 % (McGuffin *et al.*, 2000). The predicted topology of GAFA DosS differs from the core structure of a GAF domain (α-α-β-β-β-α-β-β-β-α), as its predicted secondary structure contains only 5 β-strands in the β-sheet, whereas all the structures of GAF domains contain a 6-stranded anti-parallel β-sheet. However, in general, the PSIPRED secondary structure prediction of DosS GAFA compares well to the sequence topology of GAFA obtained using PDE2A (pdb 1MC0) as a template, shown in Figure 4-15, except that PSIPRED predicts an extra α-helix after strand β-2, the absence of a β-strand (β-3) and a shift in the position of the helix-α^* by seven residues towards the N-terminus. The H149 binding region, as previously thought, is not predicted to be located

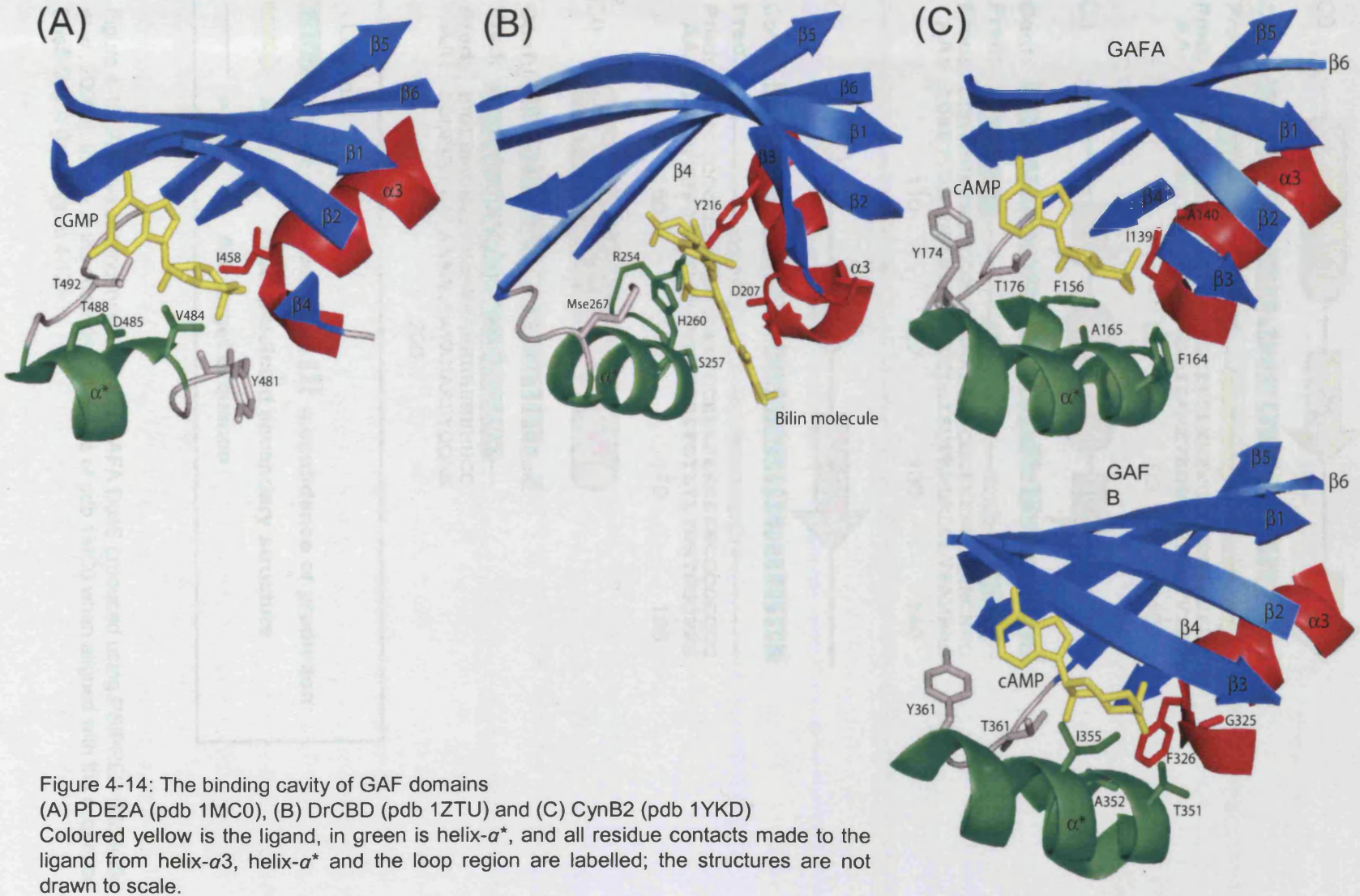


Figure 4-14: The binding cavity of GAF domains
(A) PDE2A (pdb 1MC0), (B) DrCBD (pdb 1ZTU) and (C) CynB2 (pdb 1YKD)
Coloured yellow is the ligand, in green is helix-*α**, and all residue contacts made to the ligand from helix-*α*3, helix-*α** and the loop region are labelled; the structures are not drawn to scale.

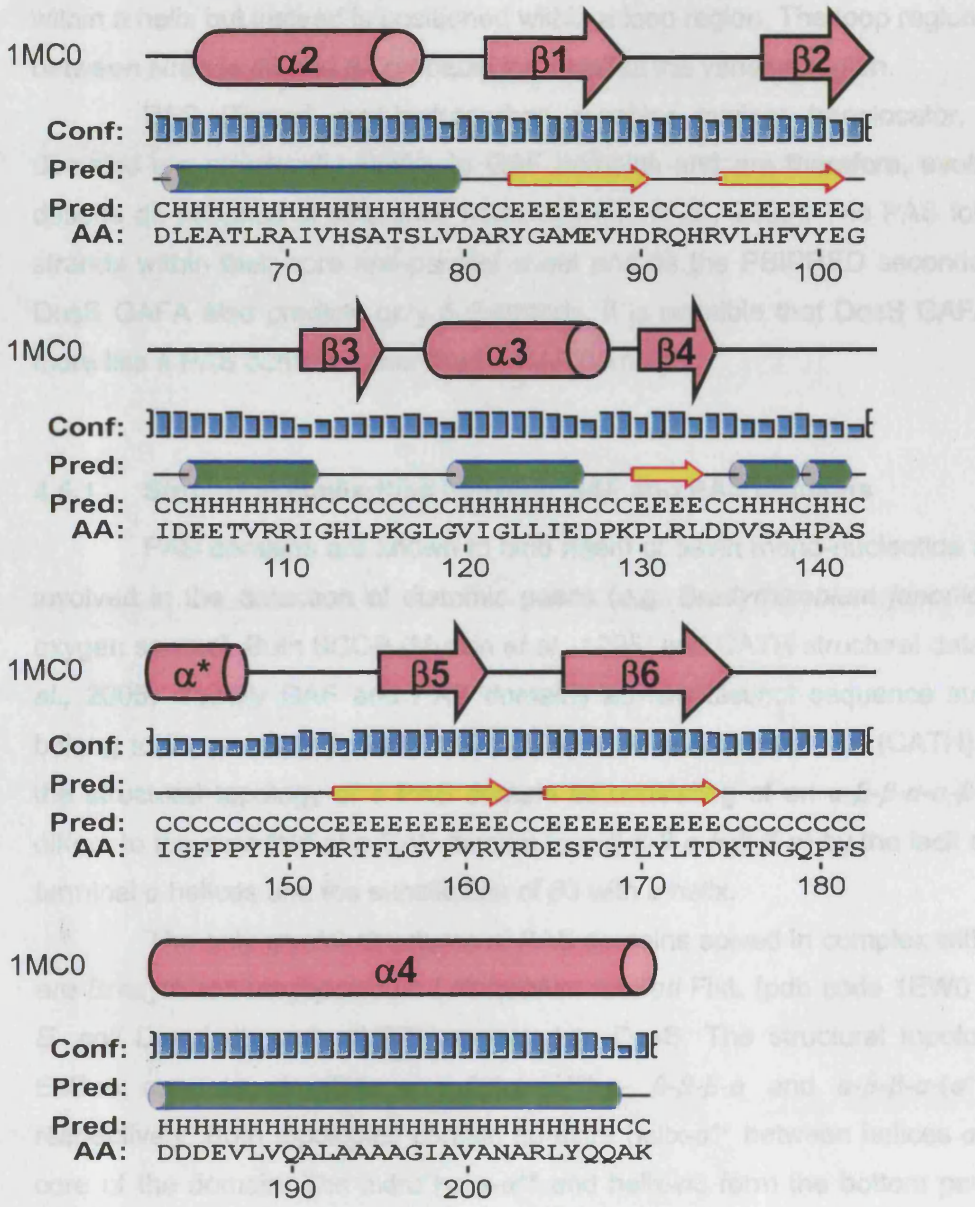


Figure 4-15: Putative secondary structure of GAFA DosS produced using PSIPRED (McGuffin *et al.*, 2000a). In pink is the secondary structure of pdb 1MC0 when aligned with the sequence DosS GAFA (from figure 4-12).

within a helix but instead is positioned within a loop region. The loop region and the α-helix between strands β3 and β4 probably form part of the variable region.

PAS (**P**eriod, **a**ryl-hydrocarbon receptor nuclear translocator, **s**ingle-minded) domains are structurally similar to GAF domains and are therefore, evolutionary related despite an absence of sequence homology (Ho *et al.*, 2000). The PAS fold contains 5 β-strands within their core anti-parallel sheet and as the PSIPRED secondary structure for DosS GAFA also predicts only 5 β-strands, it is possible that DosS GAFA is structurally more like a PAS domain rather than a GAF domain.

4.5.1 Structural similarities between GAF and PAS domains

PAS domains are known to bind haem or flavin mono-nucleotide and a sub-set is involved in the detection of diatomic gases (*e.g. Bradyrhizobium japonicum* FixL is an oxygen sensor). Both SCOP (Murzin *et al.*, 1995) and CATH structural databases (Pearl *et al.*, 2005) classify GAF and PAS domains as two distinct sequence superfamilies that belong to the same fold: Profilin-like (SCOP) or β-lactamase like (CATH). SCOP defines the structural topology of a PAS domain as consisting of an α-β-β-α-α-β-β-β fold, which differs to the core fold of a GAF domain (α-α-β-β-β-α-β-β-β-α) by the lack of the N- and C-terminal α helices and the substitution of β3 with a helix.

The only crystal structures of PAS domains solved in complex with a haem ligand are *Bradyrhizobium japonicum* / *Rhizobium meliloti* FixL (pdb code 1EW0 and 1DP9) and *E. coli* Dos (pdb code 1V9Z), unrelated to DosS. The structural topology of FixL and EcDos consists of either a α-β-β-α-(α**)-α- β-β-β-α and α-β-β-α-(α**)-α-β-β-β fold, respectively. Both topologies contain an extra helix-α** between helices α2 and α3 in the core of the domain. The extra helix-α** and helix-α3 form the bottom part of the binding cavity and the anti-parallel β-sheet forms the roof of the binding cavity (Figure 4-16). The structural formation of these elements creates a binding pocket extremely similar to that of a GAF domain. In fact, the only major, structural differences in GAF and PAS domains appear to be the number of β-strands within the β-sheet and the location of the extra helix (helix-α* or helix-α**). The anti-parallel β-sheet of PAS domains has a strand order of 2-1-5-4-3, lacking the β3 strand of a GAF domain (strand order 3-2-1-6-5-4). The secondary structure prediction of DosS GAFA using PSIPRED lacks the β3 strand (Figure 4-15) and contains an additional helix and so conforms to the topology of a PAS domain. However, in GAF domains the β3 strand is the most varied strand within the core β-sheet, for example the β3 strand of PDE2A GAFA is 10 residues in length whereas in PDE2A GAFB

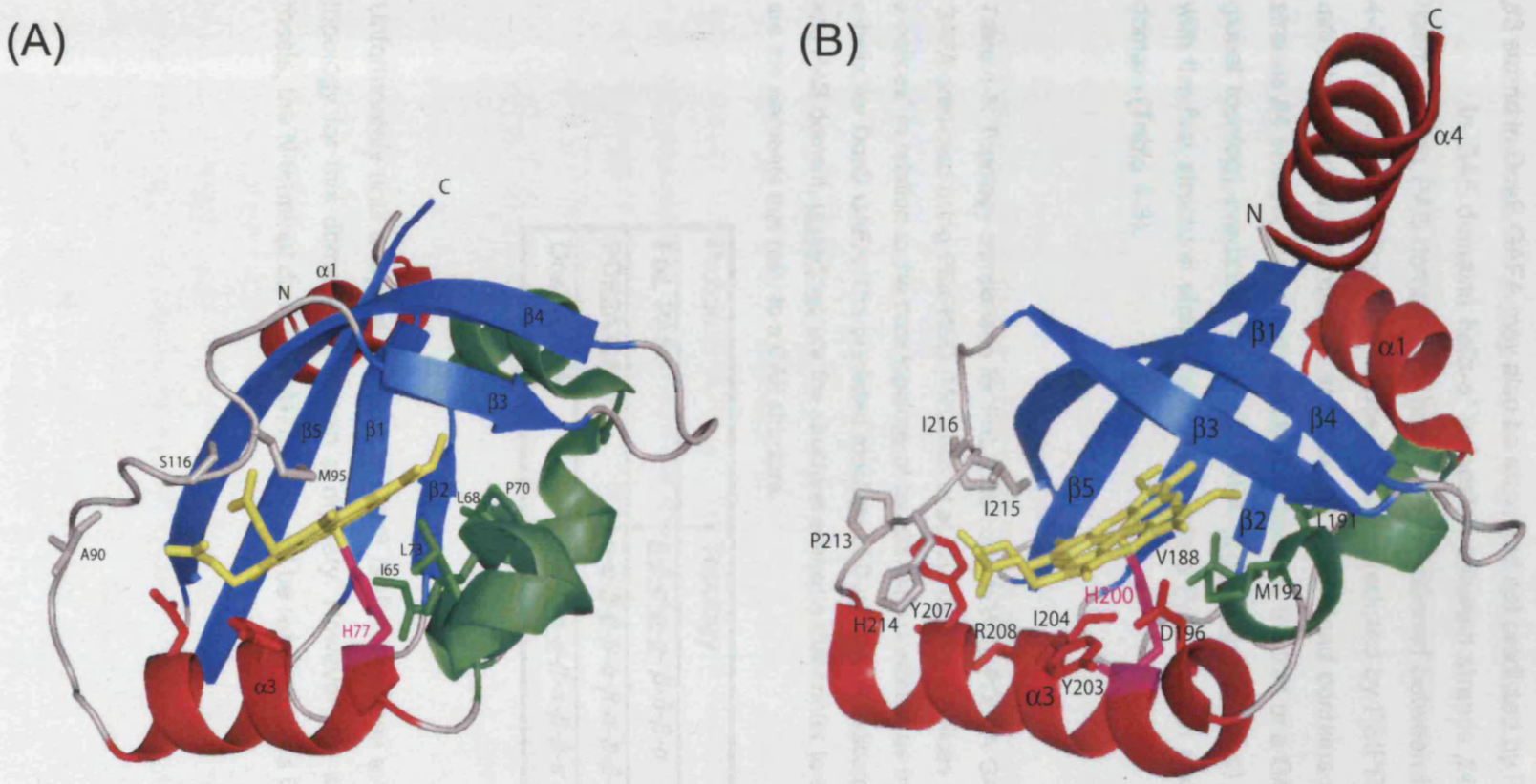


Figure 4-16: Haem binding PAS domains (A) *E. coli* Dos (pdb 1V9Z) and (B) *Rhizobium meliloti* FixL (pdb 1DP9). Shown in green is helix-α^{**}, the haem ligand is coloured in yellow and all residue contacts made to the ligand from helix-α3, helix-α^{**} and loop region are labelled; ;the structures are not drawn to scale.

it is only 4 residues long. As PDE2A GAFB is a sequence homologue of DosS GAFA the β3 strand in DosS GAFA may also be short and not predicted by PSIPRED.

In GAF domains helix-α* is located between strands β4 and β5, whereas in the haem binding PAS domains the helix-α** is positioned between helices α2 and α3 (Table 4-3). The structural topology of DosS GAFA, predicted by PSIPRED, does not contain an extra helix between strand β2 and helix α2 but instead contains an extra helix α between strands β4 and β5, suggesting similarities towards the fold of a GAF domain. However, the global topology predicted for DosS GAFA (*α-β-β-α-α-β-α-β-β-α*) appears to be a hybrid, with the first structural elements conferring to a PAS domain and the second half a *GAF* domain (Table 4-3).

Table 4-3: Topology comparison for FixL (pdb code 1EW0), PDE2A GAFB (pdb 1MC0) and DosS GAFA predicted using PSIPRED (McGuffin *et al.*, 2000). The α-helices in blue represent the "extra α-helices" in relation to the core topology of the domains predicted as by SCOP, in red is the 'extra α-helix' for DosS GAFA. The predicted topology of DosS GAFA indicates a hybrid between a GAF and PAS domain, underlined are the structural elements that confer to a PAS domain and in *italics* are the elements that refer to a GAF structure.

Protein	Topology
FixL PAS	*β-β-α-α-α- β-β-β-α*
PDE2A GAFB	*α-α-β-β-β-α-β-α-β-β-α*
DosS GAFA	*α-β-β-α-α-β-α-β-β-α*

Unfortunately until a crystal structure of DosS GAFA becomes available the real structural topology for this domain will remain a mystery. However, for simplicity, throughout this thesis, the N-terminal domain of *MTB* DosS will be referred to as GAFA.

4.6 Elucidating the mechanism of DosS function

The sensory protein DosS is part of a signal transduction system, where the histidine kinase domain of DosS undergoes autophosphorylation in response to an environmental change and transfers this signal to its cognate response regulator DosR to initiate a response (see section 1.5.3). DosS and DosR are known to be activated by hypoxia and NO. Therefore studies were undertaken to show whether autophosphorylation of DosS can occur upon binding of NO.

The kinase catalytic core of DosS refolded from inclusion bodies (residues 378-578, designated as His(6)-$DevS_{201}$) was shown to undergo autophosphorylation in the presence of [γ-^{32}P] ATP at His395 (Saini *et al.*, 2002). As the N- terminal input domain of DosS is not present to detect a stimulus and regulate the autophosphorylation process the catalytic kinase core in this construct is always activated.

Using the same domain boundaries as Saini and co-worker, the DNA coding sequence for the kinase domain of DosS was amplified by PCR using full length DosS as a template. All molecular cloning procedures were performed as described in sections 2.4.4-2.4.7 and resulted in the production of an N-terminal His-tag protein. Overexpression of the kinase catalytic domain (HK) of DosS in *E. coli* BL21 (DE3) pLysS cells and purification using affinity chromatography resulted in soluble protein (Figure 4-17), approximately 1L of cells yielded 20 mg/ml. The production of soluble protein was surprising as Saini and co-workers state that "overexpression in *E. coli* always resulted in the localization of His(6)-DevS201 in inclusion bodies", although they do not give details of alternative growth or purification conditions tried. The phosphorylation activity of soluble HK DosS domain was tested based on Mg^{2+}-dependant incorporation on ^{32}P labelled phosphate group from ATP (section 2.7.6). Figure 4-18(A) depicts the results obtained and clearly shows that the soluble HK domain of DosS is active with the protein capable of undergoing autophosphorylation as evident from incorporation of labelled phosphate into the protein in a time dependant manner. These findings are in agreement with those published previously on the refolded HK domain (Saini *et al.*, 2002).

Preliminary experiments were initiated to demonstrate a link between the binding of NO to full-length DosS and Mg^{2+} dependant autophosphorylation at His395. When the same experimental conditions for the HK domain were applied to full-length DosS, no autophosphorylation was detected. However, upon addition of diethylenetriamine-nitric oxide, a NO releasing compound, to the reaction mixture DosS undergoes autophosphorylation (Figure 4-18(B)).

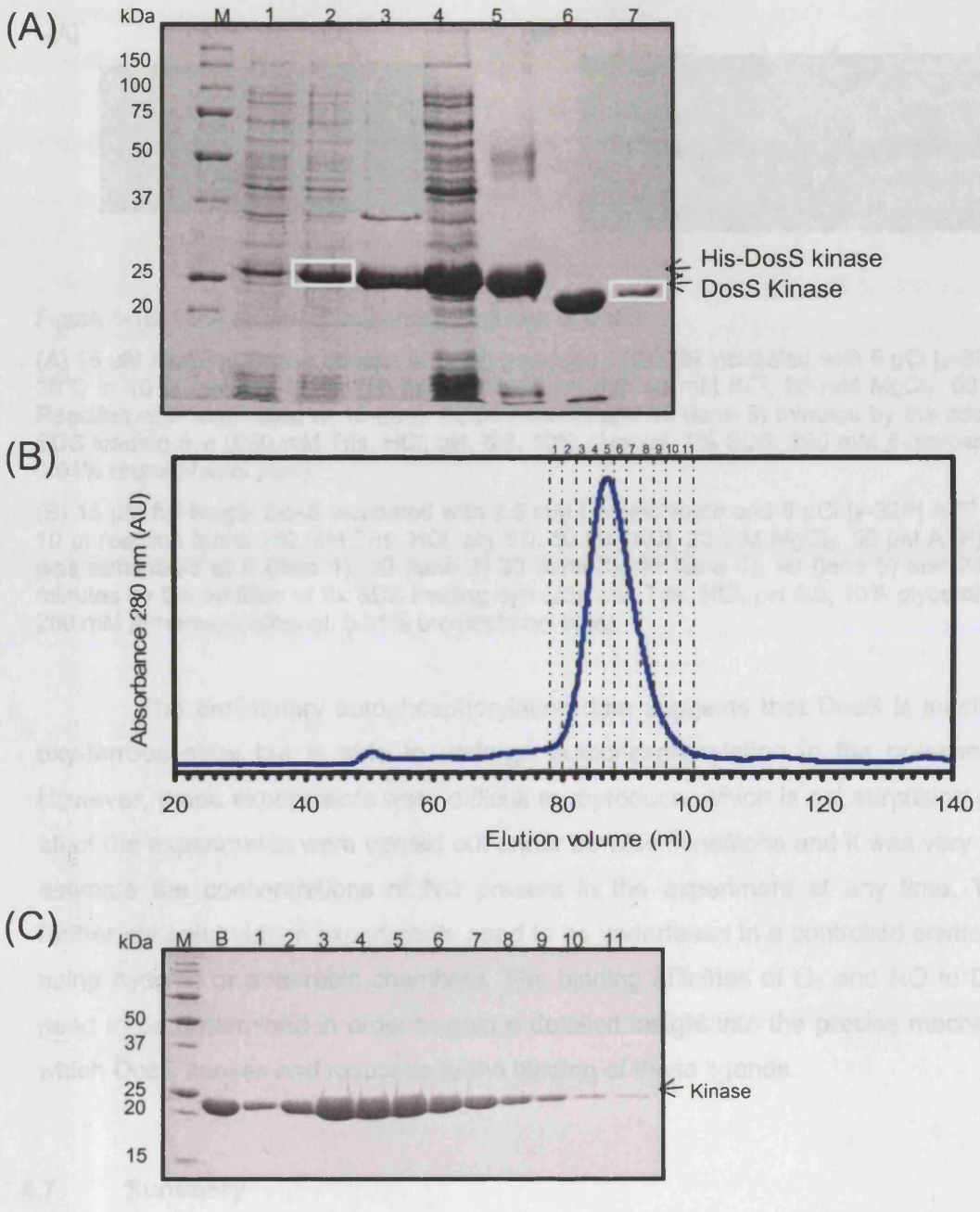


Figure 4-17: Expression and purification of DosS Kinase domain, estimated molecular weight is 22 kDa

(A) Reducing 10 % SDS-PAGE analysis of expression and purification. Lanes: M – molecular weight markers (kDa), 1 – before induction; 2 – after induction; 3 – insoluble fraction in 0.2% SDS; 4 – soluble fraction; 5 – Ni-NTA agarose elution; 6 – TEV protease cleavage; 7 – cleaved Kinase DosS domain.

(B) Size exclusion chromatography (SEC) using a Superdex 100 column, fractions collected are marked above the chromatogram.

(C) Reducing 15 % SDS-PAGE analysis of fractions from SEC (panel B). Lanes: M – molecular weight markers (kDa), other lane numbers correspond with fraction number from SEC. Fractions 1-11 from SEC were concentrated and the kinase DosS protein used in crystallisation trays and for phosphorylation experiments.

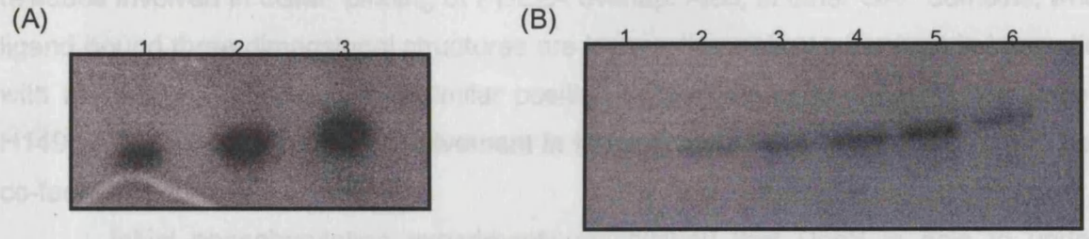


Figure 4-18: Time course of autophosphorylation of DosS:

(A) 15 µM histidine kinase domain of DosS (residues 378-578) incubated with 5 µCi [γ-32P] ATP at 30°C in 10 µl reaction buffer (50 mM Tris-HCl, pH 8.0, 50 mM KCl, 25 mM $MgCl_2$, 50 µM ATP). Reaction was terminated at 15 (lane 1), 30 (lane 2) and 45 (lane 3) minutes by the addition of 3x SDS loading dye (250 mM Tris. HCl, pH 6.8, 10% glycerol, 1% SDS, 280 mM β-mercaptoethanol, 0.01% bromophenol blue)

(B) 15 µM full length DosS incubated with 2.5 mM DetaNoNoate and 5 µCi [γ-32P] ATP at 30°C in 10 µl reaction buffer (50 mM Tris. HCl, pH 8.0, 50 mM KCl, 25 mM $MgCl_2$, 50 µM ATP). Reaction was terminated at 0 (lane 1), 10 (lane 2) 30 (lane 3), 60 (lane 4), 90 (lane 5) and 240 (lane 6) minutes by the addition of 3x SDS loading dye (250 mM Tris. HCl, pH 6.8, 10% glycerol, 1% SDS, 280 mM β-mercaptoethanol, 0.01% bromophenol blue)

The preliminary autophosphorylation data suggests that DosS is inactive in the oxy-ferrous state but is able to undergo autophosphorylation in the presence of NO. However, these experiments were difficult to reproduce, which is not surprising given that all of the experiments were carried out under aerobic conditions and it was very difficult to estimate the concentrations of NO present in the experiment at any time. Therefore, further phosphorylation experiments need to be undertaken in a controlled environment by using hypoxic or anaerobic chambers. The binding affinities of O_2 and NO to DosS also need to be determined in order to gain a detailed insight into the precise mechanisms by which DosS senses and responds to the binding of these ligands.

4.7 Summary

Visible spectroscopy of purified GAFA and full-length DosS proteins identified the presence of a classic high-spin b haem and site-directed mutagenesis identified H149 as critical to haem-binding. The reduced form of the haem co-factor is shown to bind CO, NO and O_2, consistent with the role of DosS in limited oxygen conditions. Although this is the first known GAF domain to bind haem, structural similarities between GAF and PAS domains have been previously established and PAS domains commonly use haem as a cofactor for the detection of diatomic gases. The only structural homologue identified to DosS GAFA is the GAFB domain of PDE2A. A sequence alignment shows that the topological positioning of the H149 involved in coordinating the haem iron in DosS and the

residues involved in cGMP binding of PDE2A overlap. Also, in other GAF domains, whose ligand bound three dimensional structures are known, the residues involved in interactions with the ligand are located in a similar position in the primary amino acid sequence as H149 of GAFA, supporting its involvement in forming part of the binding site for the haem co-factor.

Initial phosphorylation experiments established that DosS is able to undergo autophosphorylation in the presence of nitric oxide, providing a first glimpse into the mechanism by which DosS functions. However, further controlled experiments need to be performed in order to fully understand the molecular mechanisms by which DosS responds to hypoxia.

CHAPTER FIVE

STRUCTURAL INVESTIGATIONS OF DosS GAFB

5 Structural investigations of $DosS_{231-379}$ GAFB domain

Annotation of the DosS gene through the database search engine, SMART (section 3.1.2) identified the presence of two GAF domains, GAFA ($DosS_{63-210}$) and GAFB ($DosS_{231-379}$) within the DosS N-terminal region. Although, GAFA was identified as a haem binding domain capable of coordinating diatomic gases (Chapter 4) the role of GAFB in the signalling mechanism of DosS remains elusive. This chapter therefore describes the structural exploration of GAFB through Circular Dichroism spectroscopy, Nuclear Magnetic Resonance spectroscopy and X-ray crystallography in an attempt to identify the function of GAFB.

5.1 Structural assessment of GAFB

E. coli is the most frequently used prokaryotic host system for the over expression of recombinant proteins due to its ability to grow rapidly and to a high optical density on inexpensive substrates. However, a consequence of high level protein production in the cytoplasm of *E. coli* cells is that the bacteria often manufacture proteins that are misfolded resulting in the formation of insoluble or soluble non-functional protein aggregates. The over expression and purification of unlabelled GAFB consistently yielded 20 mg or more of purified protein per litre of media. In order to confirm that the GAFB protein overexpressed in *E. coli* is an independently folded globular domain both Circular Dichroism (CD) and Nuclear Magnetic Resonance (NMR) spectroscopy methods were used to structurally characterise the "foldedness" of GAFB.

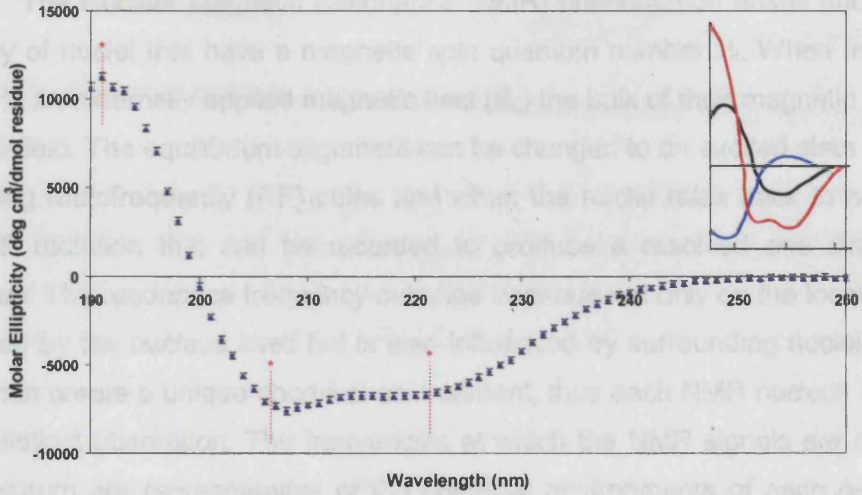
5.1.1 Circular Dichroism spectroscopy

Circular Dichroism (CD) spectroscopy is a technique that can be used to determine whether a protein is folded (section 2.7.1) and the characteristic spectra for purely α-helical, β-sheet or random coil proteins are represented in the inset of Figure 5-1. The spectra can be characterised further by estimating the proportions of secondary structure content, using software packages such as CONTIN. CONTIN uses a mathematical technique that compares the CD data generated to a dataset containing reference CD spectra for 17 proteins whose structures have been solved to high resolution by X-ray crystallography (Provencher and Glockner, 1981; Kelly and Price, 1997). The accuracy of CONTIN for determining the proportion of each secondary structure element from CD spectra was evaluated, using 17 different polypeptides whose X-ray structure had

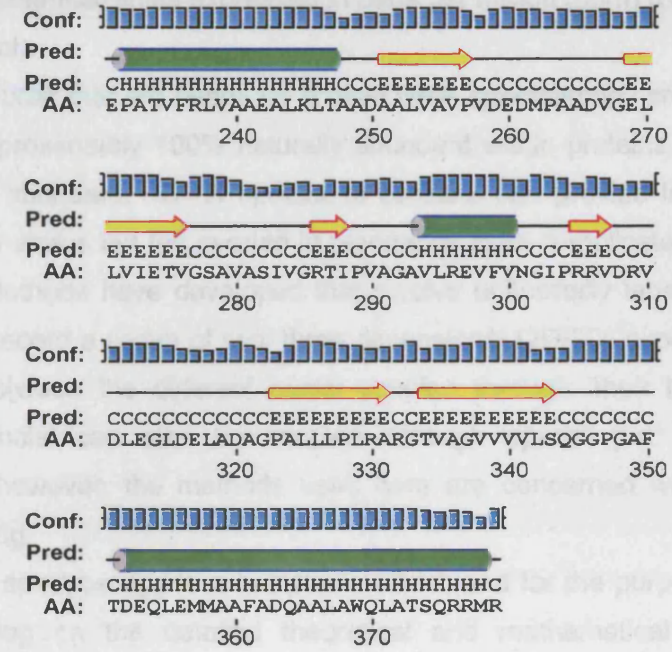
been solved (Greenfield, 1996). The correlation coefficient and mean squared error between the estimated and experimentally determined secondary structure for α-helix content was extremely accurate at 95 % whereas the accuracy of β-sheet prediction was only 60 %.

The CD signal results obtained for DosS GAFB were converted to residue molar ellipticity values (section 2.7.1) and plotted in the form of a scatter graph. The shape of DosS GAFBs CD spectrum (Figure 5-1(A)) shows two shallow negative bands at 208 nm and 222 nm and a positive band at approximately 190 nm, which is indicative of a protein containing folded globular components with a mixture of α-helices and β-sheets (inset of Figure 5-1(A)). Deconvolution of the residue molar ellipticity values using the program CONTIN predicts that the protein is composed of 48 % β-sheet, 19 % α-helix and 33 % random coil. The output from CONTIN was compared to the predicted secondary structure of GAFB produced using PSIPRED (Figure 5-1(B)), a program that uses sequence profile alignments in order to determine putative β-sheets and α-helices to an accuracy of 77 % (McGuffin *et al.*, 2000b). The secondary structure derived using PSIPRED contained a β-sheet, α-helix and random coil content of 27 %, 34 % and 39 % respectively, which is apparently in disagreement with the results obtained by CD spectral analysis using CONTIN. As the accuracy of the estimation of helical content by CONTIN is high (95 %), the CD data suggests that the number of α-helices in GAFB is lower than that predicted by PSIPRED.

(A)



(B)



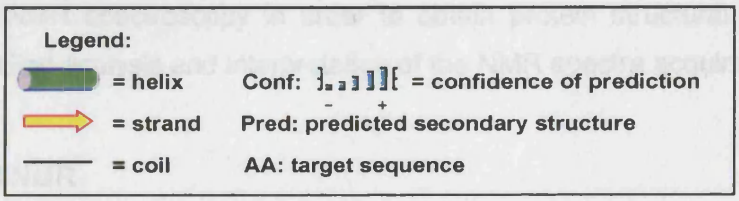


Figure 5-1: Structural analysis of DosS GAFB

(A) Circular Dichroism spectra of 24 μM GAFB recorded in 20 mM phosphate pH 7.0 using a 0.1 cm path length quartz cuvette, the inset shows representation of purely *α*-helical, *β*-sheet and random coil spectra.

(B) Putative secondary structure of GAFB DosS predicted using PsiPred (McGuffin *et al.*, 2000a).

5.1.2 Nuclear Magnetic Resonance

The Nuclear Magnetic Resonance (NMR) phenomenon arises due to an intrinsic property of nuclei that have a magnetic spin quantum number ½. When these nuclei are placed in an externally applied magnetic field (B_0) the bulk of their magnetic moments align with the field. The equilibrium alignment can be changed to an excited state by applying an oscillating radiofrequency (RF) pulse and when the nuclei relax back to equilibrium they emit RF radiation that can be recorded to produce a resolved one dimensional (1D) spectrum. The resonance frequency detected depends not only on the local magnetic field produced by the nucleus itself but is also influenced by surrounding nuclei. Neighbouring nuclei can create a unique chemical environment, thus each NMR nucleus of a given type has a distinct absorption. The frequencies at which the NMR signals are recorded in the 1D spectrum are representative of the chemical environments of each nucleus and are referred to as chemical shifts expressed in parts per million (ppm) (determined relative to a reference signal).

The nuclei that are useful for protein NMR spectroscopy are ^{1}H, ^{13}C and ^{15}N, and while ^{1}H is approximately 100% naturally abundant within proteins, both ^{13}C and ^{15}N are less than 1% abundant. 1D ^{1}H spectra of proteins can provide information about their folded globular status but the overlap in resonance lines complicates further interpretation of the data. Methods have developed that involve uniformally labelling proteins with ^{15}N or/and ^{13}C to record a series of two/ three dimensional (2D/3D) experiments that correlate information between the different nuclei coupled through 'their bonds' (*i.e.* via scalar couplings). Nuclei can also be coupled 'through space' (*i.e.* through dipole-dipole interactions), however, the methods used here are concerned with the former type of nuclear coupling.

NMR spectroscopy is a complex method, and for the purpose of this thesis I will not be focusing on the detailed theoretical and mathematical aspects but on the application of NMR spectroscopy in order to obtain protein structural information. This includes a detailed analysis and interpretation of the NMR spectra acquired.

5.2 1D ^{1}H NMR

The folded status of purified protein can be assessed using 1D ^{1}H NMR spectroscopy. For a flexible, unfolded polypeptide chain the frequencies at which the ^{1}H NMR resonances appear, the chemical shift values are dependant on the chemical group to which the proton is attached. Every chemically equivalent group, for example a valine methyl group, will have a signal at a similar frequency. In the presence of a globular folded

protein the resonance lines not only depend on the bonded electrons but also the non-bonded electrons of the atoms that are physically close to in space. These neighbouring atoms create a unique electronic environment and so the signals for the valine methyl groups are no longer typically equivalent but dispersed over a range of frequencies. The neighbouring electronic cloud can either 'shield' or 'deshield' a nucleus from the applied magnetic field (B_0). If the nuclei are shielded the resonance frequency is lower and the signal shifts to the right of the spectrum, whereas if a deshielding effect takes place the resonance frequency is higher and the peak appears to the left of the spectrum. In a flexible, unfolded protein, the methyl and backbone amide protons display typical chemical shift values of between 0.8-1.2 ppm and 7.0-8.5 ppm, respectively (Figure 5-2(A)). Signals to the right of 0.8 ppm indicate the presence of methyl groups that are shielded from B_0, for example being close to the delocalised ring of an aromatic group. Signals to the left of 8.5 ppm indicate the presence of backbone amide protons that are deshielded from B_0, for example an NH hydrogen bonded to a carbonyl group. Any ^1H chemical shift dispersion above 8.5 ppm and below 0.8 ppm is indicative of a wide variety of different structural regions within the tertiary structure of a folded, globular protein; these dispersions are not observed in unfolded or randomly coiled polypeptide chains (Figure 5-2(A)).

The 1D ^1H NMR spectrum of an unlabelled 0.2 mM GAFB sample was obtained at 25 °C on a Varian UnityPLUS spectrometer, operating at a nominal ^1H frequency of 500 MHz. The 1D ^1H spectrum of GAFB (Figure 5-2(B)) shows a high level of ^1H chemical shift dispersion in the methyl proton (-0.5 to 1.5 ppm) and amide proton (6-11 ppm) regions of the spectrum and indicates the presence of an independently folded globular domain. For comparative purposes a 1D ^1H spectrum of an unfolded protein is shown in Figure 5-2(A), and shows the clustering of the methyl and amide resonances around 7.5-8.5 ppm and 0.6-1.2 ppm respectively. The resonance lines observed within the 1D ^1H NMR spectrum of GAFB are sharp, indicative that GAFB is a well behaved, non-aggregating protein at micro-molar concentrations.

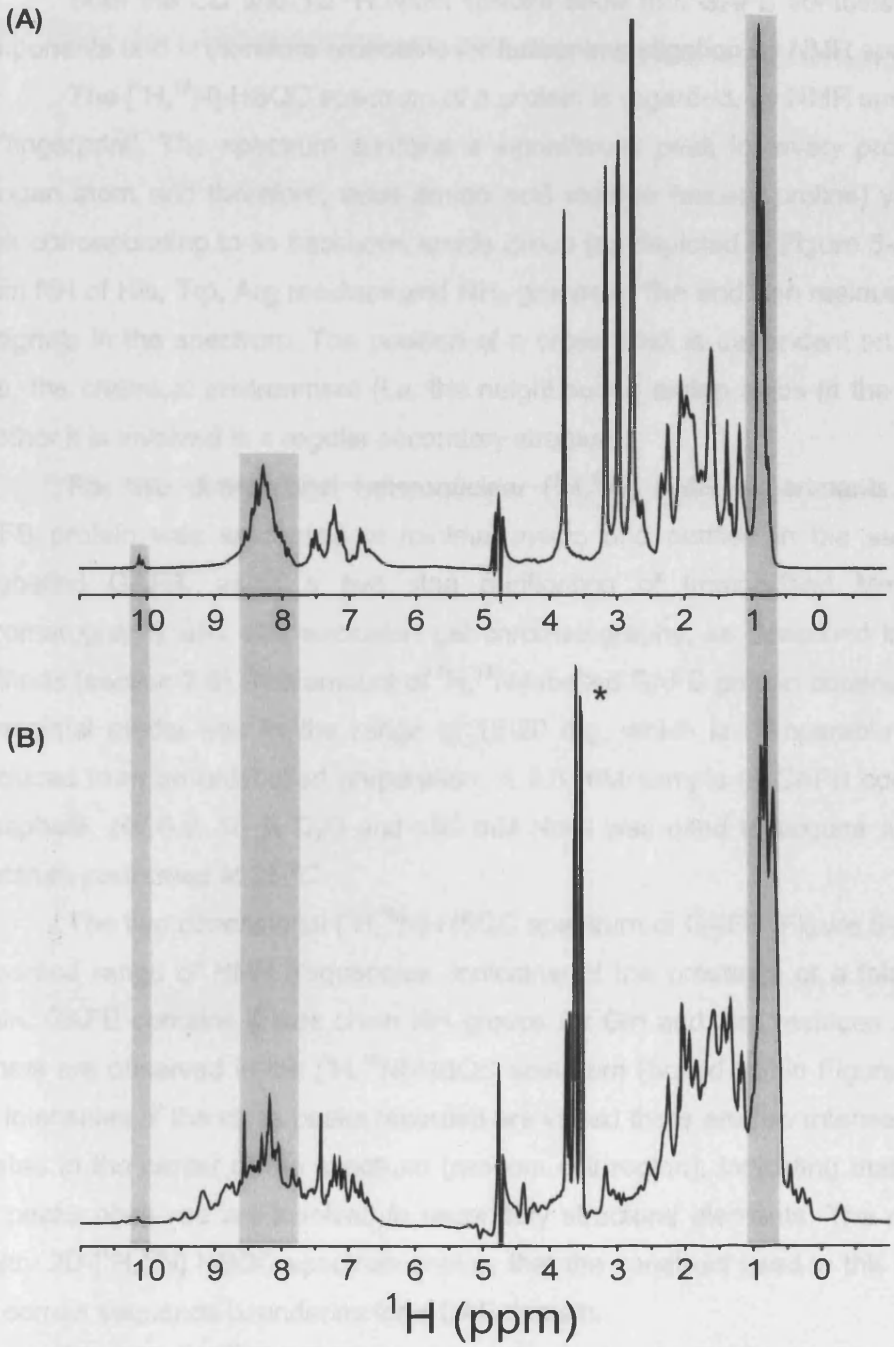


Figure 5-2: 1D ^{1}H NMR spectra

(A) HAX1 (an unfolded protein).

(B) GAFB in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 and 10 % glycerol

The grey boxes illustrate the random coil regions for methyl group resonances (~0.8 ppm), NH groups (7.5-8.5 ppm) and tryptophan side-chain NH signals (~10.2 ppm). The * represent glycerol.

5.3 [^{1}H,^{15}N] 2D Heteronuclear single quantum correlation spectroscopy (HSQC) of GAFB

Both the CD and 1D ^{1}H NMR spectra show that GAFB contains folded globular components and is therefore amenable for further investigation by NMR spectroscopy.

The [^{1}H,^{15}N]-HSQC spectrum of a protein is regarded, by NMR spectroscopists as its 'fingerprint'. The spectrum contains a signal/cross peak for every proton bound to a nitrogen atom and therefore, each amino acid residue (except proline) yields one cross peak corresponding to its backbone amide group (as depicted in Figure 5-3(A)). The side-chain NH of His, Trp, Arg residues and NH_2 groups of Gln and Asn residues also give rise to signals in the spectrum. The position of a cross peak is dependent on the amino acid type, the chemical environment (*i.e.* the neighbouring amino acids in the sequence) and whether it is involved in a regular secondary structure.

For two dimensional heteronuclear [^{1}H,^{15}N] NMR experiments ^{1}H,^{15}N-labelled GAFB protein was expressed in minimal media and purified in the same manner as unlabelled GAFB, using a two step purification of Immobilised Metal Ion Affinity Chromatography and size exclusion gel chromatography, as described in Materials and Methods (section 2.6). The amount of ^{1}H,^{15}N-labelled GAFB protein obtained from one litre of minimal media was in the range of 15-20 mg, which is comparable to the amount produced from an unlabelled preparation. A 0.6 mM sample of GAFB containing 20 mM phosphate, pH 6.0, 10 % D_2O and 100 mM NaCl was used to acquire a [^{1}H,^{15}N]-HSQC spectrum performed at 25 °C.

The two dimensional [^{1}H,^{15}N]-HSQC spectrum of GAFB (Figure 5-3(B)) displays a dispersed range of NMR frequencies, indicative of the presence of a folded polypeptide chain. GAFB contains 6 side chain NH groups for Gln and Asn residues and all of these signals are observed in the [^{1}H,^{15}N]-HSQC spectrum (boxed within Figure 5-3(B)). Whilst the intensities of the cross peaks recorded are varied there are few intense or sharp peaks located in the center of the spectrum (random coil region), indicating that the majority of the peaks observed are involved in secondary structural elements. The result of a good quality 2D-[^{1}H,^{15}N] HSQC spectrum implies that the construct used in this experiment has the correct sequence boundaries for a GAF domain.

The 2D-[^{1}H,^{15}N] HSQC spectra of proteins in itself, provides little structural information about the protein unless the assignment of the individual cross peaks to a specific amino acid is obtained. Therefore, to enable sequential assignment of the cross peaks, a set of triple resonance experiments were acquired on a double ^{15}N,^{13}C-labelled sample of GAFB.

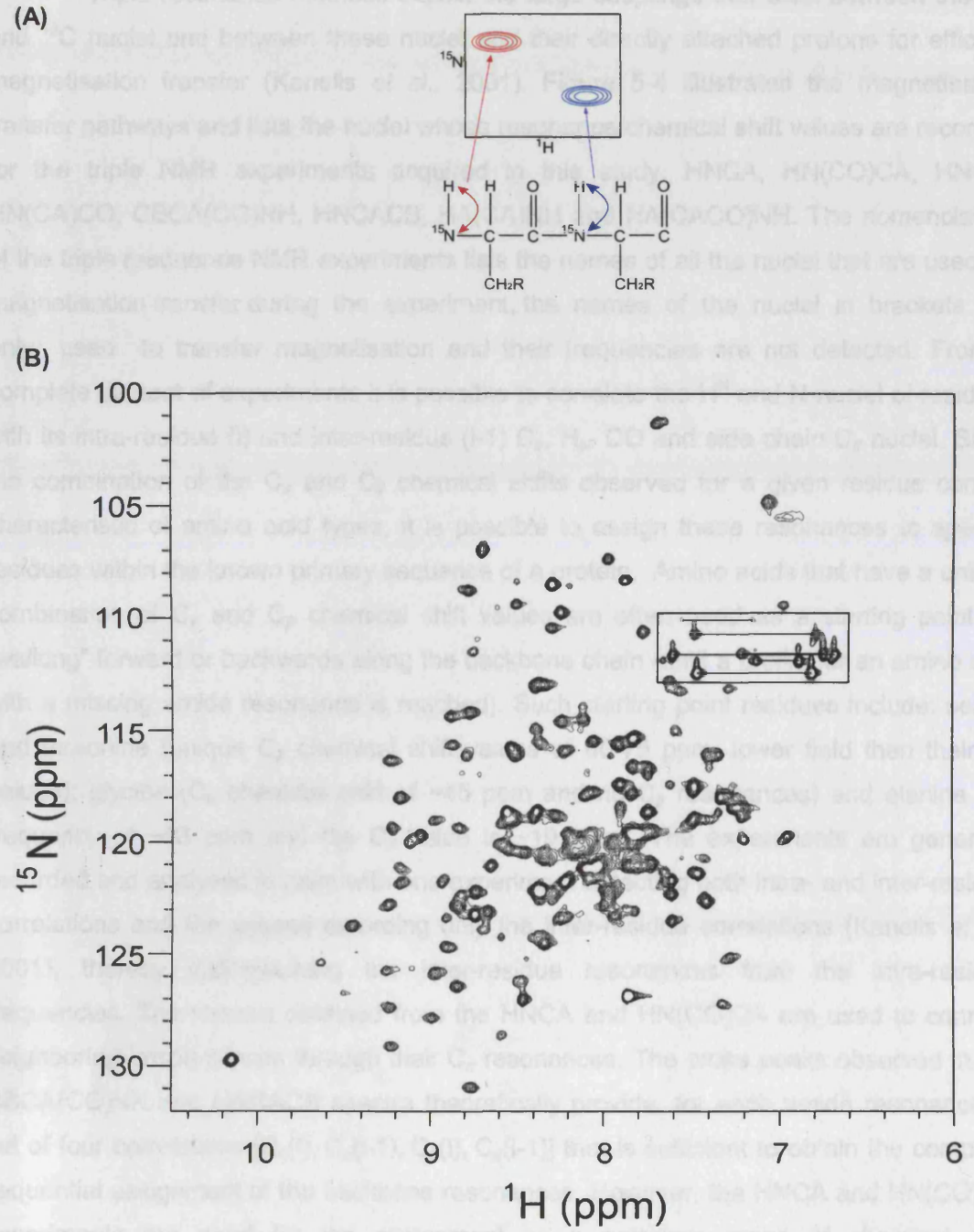


Figure 5-3: 2D-[^{1}H,^{15}N] Heteronuclear single quantum correlation (HSQC) spectroscopy

(A) Figure shows how cross peaks within the [^{1}H,^{15}N]-HSQC are produced in reference to a ^{15}N polypeptide chain.

(B) 2D [^{1}H,^{15}N]-HSQC spectrum of [^{15}N]-labelled GAFB in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0; recorded on a 600 MHz Varian Inova spectrometer at 298K. Boxed are pairs of Gln/Asn side chain NH_2 cross peaks connected by horizontal lines.

5.4 Triple resonance NMR experiments and backbone assignments of GAFB

Triple resonance methods exploit the large couplings that exist between the ^{15}N and ^{13}C nuclei and between these nuclei and their directly attached protons for efficient magnetisation transfer (Kanelis *et al.*, 2001). Figure 5-4 illustrates the magnetisation transfer pathways and lists the nuclei whose resonance/chemical shift values are recorded for the triple NMR experiments acquired in this study, HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, HA(CA)NH and HA(CACO)NH. The nomenclature of the triple resonance NMR experiments lists the names of all the nuclei that are used for magnetisation transfer during the experiment, the names of the nuclei in brackets are only used to transfer magnetisation and their frequencies are not detected. From a complete dataset of experiments it is possible to correlate the H^N and N nuclei of residue i with its intra-residue (i) and inter-residue (i-1) C_α, H_α, CO and side chain C_β nuclei. Since the combination of the C_α and C_β chemical shifts observed for a given residue can be characteristic of amino acid types, it is possible to assign these resonances to specific residues within the known primary sequence of a protein. Amino acids that have a unique combination of C_α and C_β chemical shift values are often used as a starting point for "walking" forward or backwards along the backbone chain (until a proline or an amino acid with a missing amide resonance is reached). Such starting point residues include: serine and threonine (unique C_β chemical shift values of 60-75 ppm, lower field than their C_α values); glycine (C_α chemical shift of ~45 ppm and no C_β resonances) and alanine (C_α frequency of ~53 ppm and the C_β value is ~19 ppm). The experiments are generally recorded and analysed in pairs with one experiment detecting both intra- and inter-residue correlations and the second recording only the inter-residue correlations (Kanelis *et al.*, 2001), thereby distinguishing the inter-residue resonances from the intra-residue frequencies. The spectra obtained from the HNCA and HN(CO)CA are used to connect neighboring amide groups through their C_α resonances. The cross peaks observed in the CBCA(CO)NH and HNCACB spectra theoretically provide, for each amide resonance, a set of four correlations [C_α(i), C_α(i-1), C_β(i), C_β(i-1)] that is sufficient to obtain the complete sequential assignment of the backbone resonances. However, the HNCA and HN(CO)CA experiments are used for the assignment as a narrower range of chemical shift frequencies (70-45 ppm for HNCA, rather than 70-15 ppm for the HNCABA) are recorded with increased resolution of C_α region and therefore decreasing potential overlap of cross peaks. The primary sequence of DosS GAFB is highly degenerate, as it is composed of many similar amino acids; for example Alanines, Leucines and Valines form 19 %, 12 %

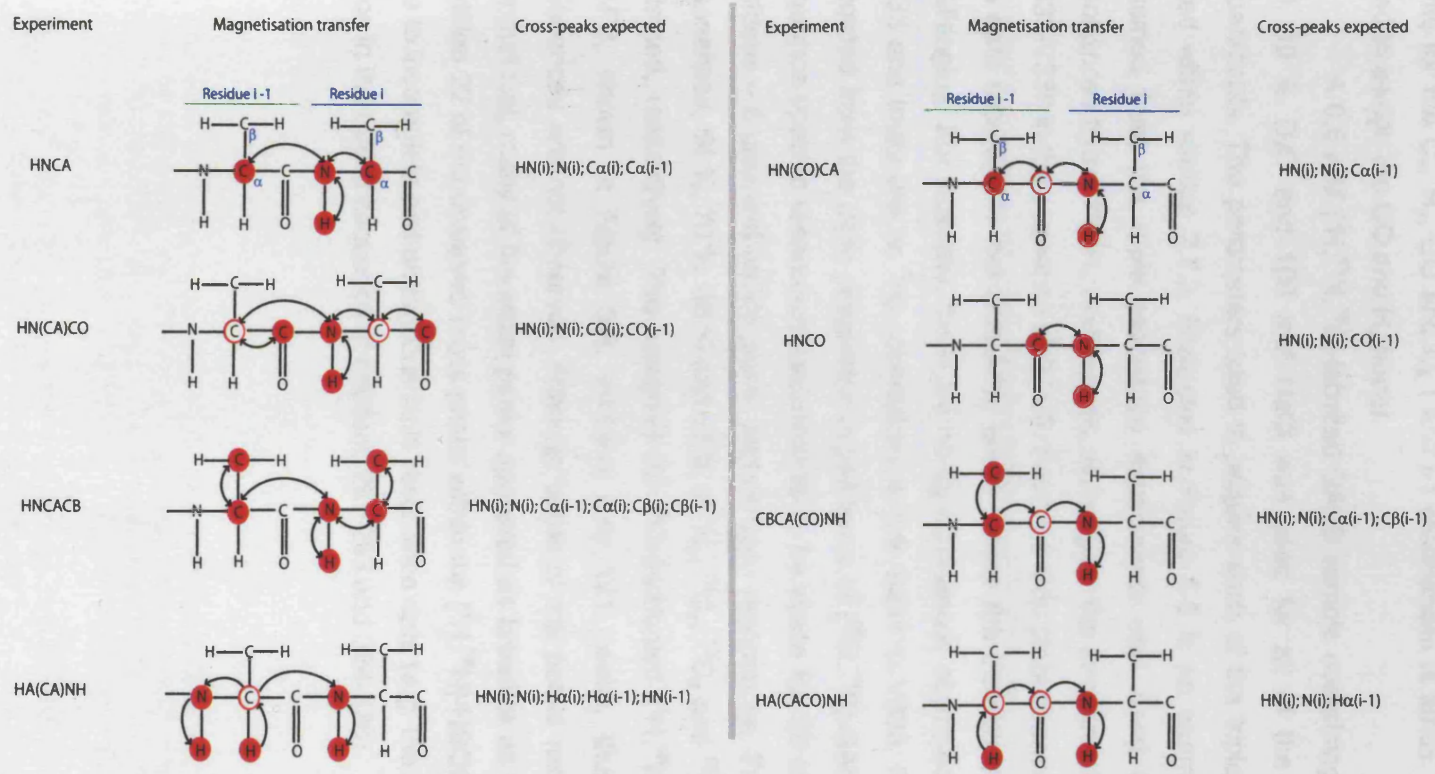


Figure 5-4: Experimental strategy for the resonance assignment of backbone $^{1}H^{N}$, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{1}H_{\alpha}$ and CO nuclei in a dipeptide fragment showing residues i and i-1. The direction of the magnetisation transfer pathways in the HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCA(CO)NH HNCACB, HA(CA)NH and HA(CACO)NH experiments are indicated by arrows. Spins whose chemical shift are recorded are shown in red circles and spins whose chemical shifts are not involved are put in parentheses in the pulse sequence name and outlined in a red circle.

and 14 %, respectively, of its amino acid sequence. For this reason, the triple resonance experiment pairs HNCO/HN(CA)CO and HA(CA)NH/HA(CACO)NH were also recorded, to overcome the ambiguity as the likelihood of two amides having identical chemical shifts for the C_α, H_α, CO and C_β i and i-1 resonances is small. These spectra are also used to assign the CO and H_α nuclei.

A 0.6 mM [^{1}H,^{15}N,^{13}C]-labelled GAFB sample containing 20 mM phosphate, pH 6.0, 10 % D_2O and 100 mM NaCl was used for all of the NMR triple resonance experiments. The parameters used to acquire each of the triple resonance spectra are listed within section 2.7.3. Illustrated in Figure 5-5 is an example of how the dataset obtained from the triple resonance experiments was used to correlate the ^{1}H/^{15}N resonances to the ^{13}C$^1H_\alpha$ resonances, allowing for the sequential assignment of residues G333-V338 in the sequence of GAFB. From the strip plots shown, it is clearly evident that the data recorded is not complete, which makes the backbone assignment technically challenging. For example, there are no C_β resonances observed for residues V335 and V338 and there are no $^1H_\alpha$ correlations to link G337 to V338. Full analysis of the data recorded from the triple resonance experiments of [^{15}N,^{13}C]-GAFB allowed for definitive sequence-specific resonance assignments to be made for 99 out of 141 residues (149 residues – 8 prolines) amide proton and nitrogen resonances. From the assigned amide resonances, 64 %, 70 %, 66 % and 63 % of $^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$ and ^{13}CO chemical shifts were obtained, respectively. The assigned 2D ^{13}C-decoupled [^{1}H,^{15}N]-HSQC spectrum for GAFB, shown in Figure 5-6, contains only 121 peaks, thus 20 backbone amide resonances are not observed. Although some of the peaks recorded were strong and symmetrical, many of the cross peaks appeared as broad or as multiple cross peaks. In addition 22 of the observed cross peaks within the [^{1}H,^{15}N]-HSQC could not be assigned due to incomplete and ambiguous triple resonance data (*e.g.* the sequence AAL, appears twice in the GAFB sequence at residues 251-253 and 324-326).

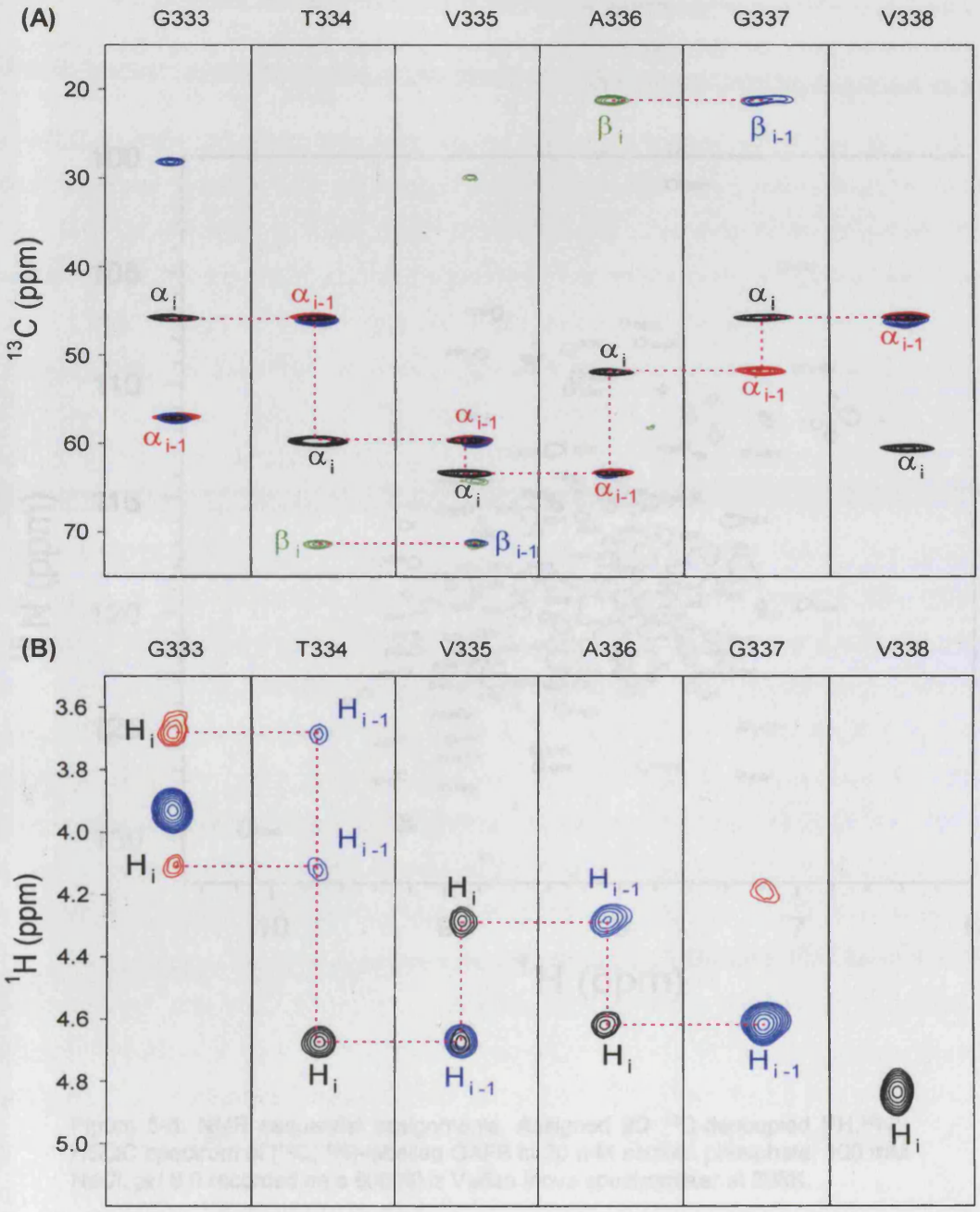


Figure 5-5: Strip plots showing correlations from residues G333 to V338 for:

(A) HNCA (black), HN(CO)CA (red), HNCACB (green, only CB correlations shown) and CBCA(CO)NH (blue) spectra showing examples of $^{1}H^{N}/^{15}N \rightarrow {}^{13}C$ correlations. Cross peaks are labelled as originating from α or β, intra- (i) and inter-residue (i-1) carbon atom correlations.

(B) HA(CA)NH (black, positive contours, and red, negative contours) and HA(CACO)NH (blue) showing the $^{1}H^{N}/^{15}N \rightarrow {}^{1}H\alpha$ correlations.

The dotted magenta line depicts the sequential assignment of selected $^{1}H^{N}/^{15}N \rightarrow {}^{13}C\alpha_i, {}^{13}C\alpha_{i-1}, {}^{1}H\alpha_i$ and $^{1}H\alpha_{i-1}$ correlations.

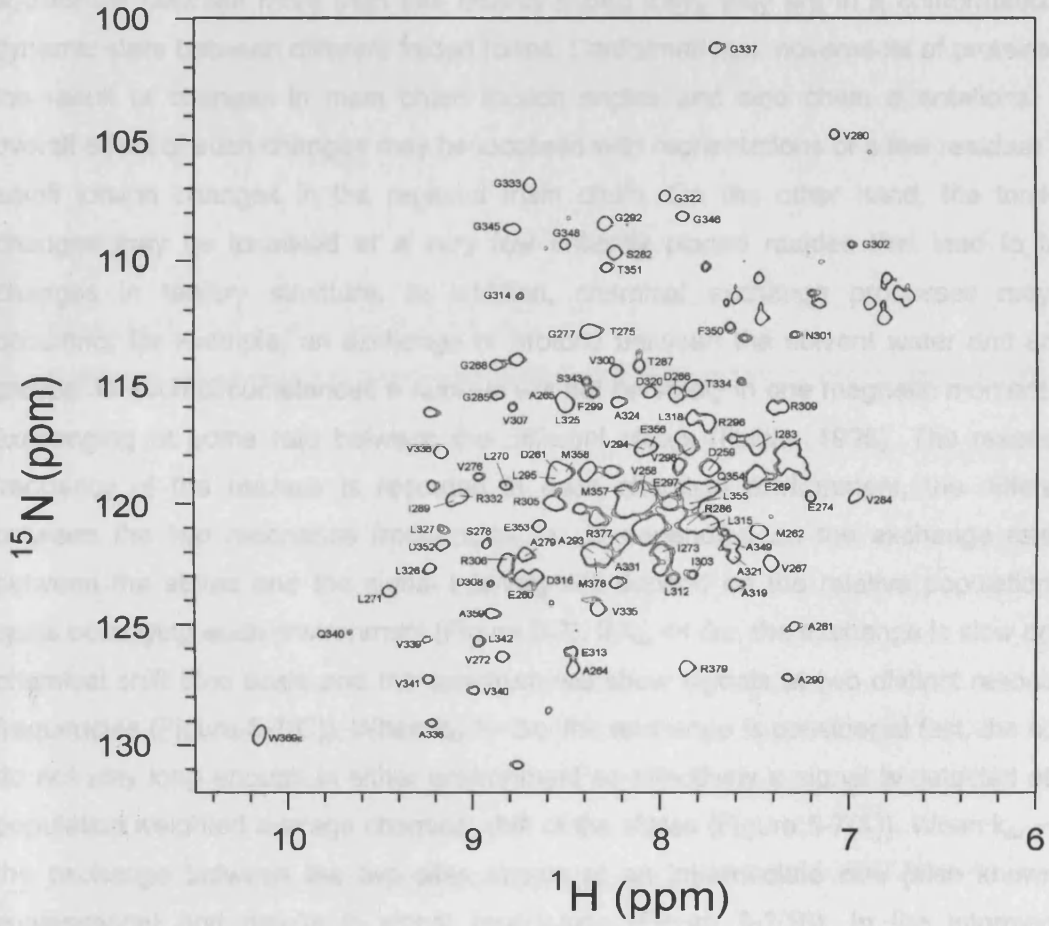


Figure 5-6: NMR sequential assignments. Assigned 2D ^{13}C-decoupled [^{1}H,^{15}N]-HSQC spectrum of [^{13}C,^{15}N]-labelled GAFB in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 recorded on a 600 MHz Varian Inova spectrometer at 298K.

In general, independently folded globular domains adopt a predominantly well defined conformation where each of the backbone amide groups in the [^{1}H,^{15}N]-HQSC spectrum will have a unique chemical shift in the proton and nitrogen dimensions and appear as dispersed peaks within the spectrum. However, some proteins exist in equilibrium between more than one distinct folded form, they are in a conformationally dynamic state between different folded forms. Conformational movements of proteins are the result of changes in main chain torsion angles and side chain orientations. The overall effect of such changes may be localised with reorientations of a few residues and small torsion changes in the regional main chain. On the other hand, the torsional changes may be localised at a very few critically placed resides that lead to large changes in tertiary structure. In addition, chemical exchange processes may be occurring, for example, an exchange of protons between the solvent water and amino groups. In such circumstances a nucleus will not be solely in one magnetic moment, but exchanging at some rate between the different states (Rattle, 1995). The resonance frequency of the nucleus is recorded in each chemical environment, the difference between the two resonance frequencies $\Delta\upsilon$, is dependent on the exchange rate k_{ex} between the states and the signal intensity will depend on the relative populations of spins occupying each environment (Figure 5-7). If $k_{ex} << \Delta\upsilon$, the exchange is slow on the chemical shift time scale and the spectrum will show signals at two distinct resonance frequencies (Figure 5-7(C)). When $k_{ex} >> \Delta\upsilon$, the exchange is considered fast, the nuclei do not stay long enough in either environment so effectively a signal is detected at the population weighted average chemical shift of the states (Figure 5-7(A)). When $k_{ex} \sim \Delta\upsilon$, the exchange between the two sites occurs at an intermediate rate (also known as coalescence) and results in signal broadening (Figure 5-7(B)). In the intermediate exchange regime, if $\Delta\upsilon$ is large enough it may cause peaks in the spectrum to disappear because the broadening becomes so great that the intensity of the resonance line becomes indistinguishable from the baseline noise (Cavanagh *et al.*, 1996).

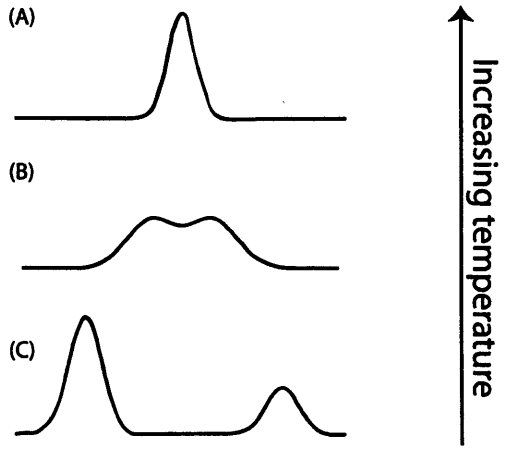


Figure 5-7: Effects of exchange between two environments. A – FAST EXCHANGE; B – INTERMEDIATE EXCHANGE and C SLOW EXCHANGE.

The variable peak appearance, missing peaks and the presence of multiple cross peaks for a single amide observed in the [^{1}H,^{15}N]-HSQC spectrum of GAFB suggests that a subset of the nuclei are experiencing chemical or conformational exchange processes on an intermediate and slow timescale. The majority of the amino acids not assigned, whether the cross peaks are missing or could not be assigned unambiguously, are located either at the start of the N-terminal (residues 231-250) or at the end of the C-terminal (residues 361-377) portion of the protein (Figure 5-8(A)). Thus it could be postulated that those regions of GAFB are undergoing conformational exchange.

To date, there are no reported three dimensional structures of GAF domains solved by NMR spectroscopy. However, the backbone ^{1}H, ^{13}C and ^{15}N resonance assignments of a GAFA domain from Phosphodiesterase 5 (PDE5) have been reported (Sekharan *et al.,* 2005). The PDE5 GAFA construct contains a total of 189 amino acids; excluding proline residues, the N-terminal residue and the C-terminal 6-histidine-tag and linker residues (Figure 5-8(B)). The [^{1}H,^{15}N]-HSQC spectrum acquired on this domain at 37°C, contained a total of 152 distinct peaks, of which 117 have been assigned, the remaining observed peaks were not assigned due to peak overlap. The majority of the unassigned residues (whether through cross peak overlap or cross peaks not present in the spectrum) are also located at the N- and C- termini portions of the protein (Figure 5-8(B)) corresponding to regions of predicted α-helical secondary structure (Sopory *et al.,* 2003).

(A) EPATVFRLVA AEALKLTAAD AALVAVPVDE DMPAADVGEL
LVIETVGSAV ASIVGRTIPV AGAVLREVFV NGIPRRVDRV
DLEGLDELAD AGPALLLPLR ARGTVAGVVV VLSQGGPGAF
TDEQLEMMAA FADQAALAWQ LATSQRRMR

(B) PLTPPRFDSD EVDQCSRLLE LVKDISSHLD VTALCHKIFL
HIHGLISADR YSLFLVCEDS SKDKFLISRL FDVAEGSTLE
EASNNCIRLE WNKGIVGHVA AFGEPLNIKD AYEDPRFNAE
VDQITGYKTQ SILCMPIKNH REEVVGVAQA INKKSGNGGT
FTEKDEKDFA AYLAFCGIVL HNAQLYETSL LENKRN

Figure 5-8: the amino acid sequence of (A) $DosS_{231-379}$ GAFB and (B) $PDE5_{125-320}$ GAFA (Sekharan *et al.*, 2005). Residues coloured red are those where assignments have been made, residues highlighted green are those that have been tentatively assigned and proline residues are coloured blue.

5.5 Identification of secondary structure elements from chemical shift data

It is well established that nuclei, which are part of secondary structural elements, show specific chemical shift changes in relation to their random coil conformational values. The chemical shifts of C_α and carbonyl carbon nuclei experience an upfield shift when present in β-sheets and a downfield shift when present in α-helices. Conversely, C_β and $^1H_\alpha$ nuclei chemical shifts have the opposite correlation with local secondary structure (Spera and Bax, 1991). Therefore, the experimental $^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$ and CO chemical shifts obtained from the GAFB assignment data were subtracted from the tabulated residue-specific random coil values (Wishart *et al.*, 1995) to yield ΔC_α, ΔC_β, ΔH_α and ΔCO.

Since the deviations ΔC_α and ΔC_β are of similar magnitude and opposite signs whether in an α-helix or β-sheet, the difference between the two deviations (ΔC_α -ΔC_β) enhances the correlation between the definition of the secondary structure and the chemical shift deviations. Where the carbon shifts were available, ΔC_α -ΔC_β values were then calculated for each GAFB residue and the (ΔC_α -ΔC_β), -ΔH_α and ΔCO data smoothed by averaging the values over three consecutive residues (Salzmann *et al.*, 2000). Figure 5-9(A), (B) and (C) shows a plot of $(\Delta C_\alpha - \Delta C_\beta)_{smoothed}$, $(-\Delta H_\alpha)_{smoothed}$ and $(\Delta CO)_{smoothed}$ chemical shift values as a function of residue number. For clarity the $(-\Delta H_\alpha)_{smoothed}$ is plotted so that the trend of positive deviation from random coil chemical shift is due to β-strands and a negative deviation for α-helices. Therefore all positive values are an indication of β-sheet elements whereas negative values indicate the possible presence of α- helices.

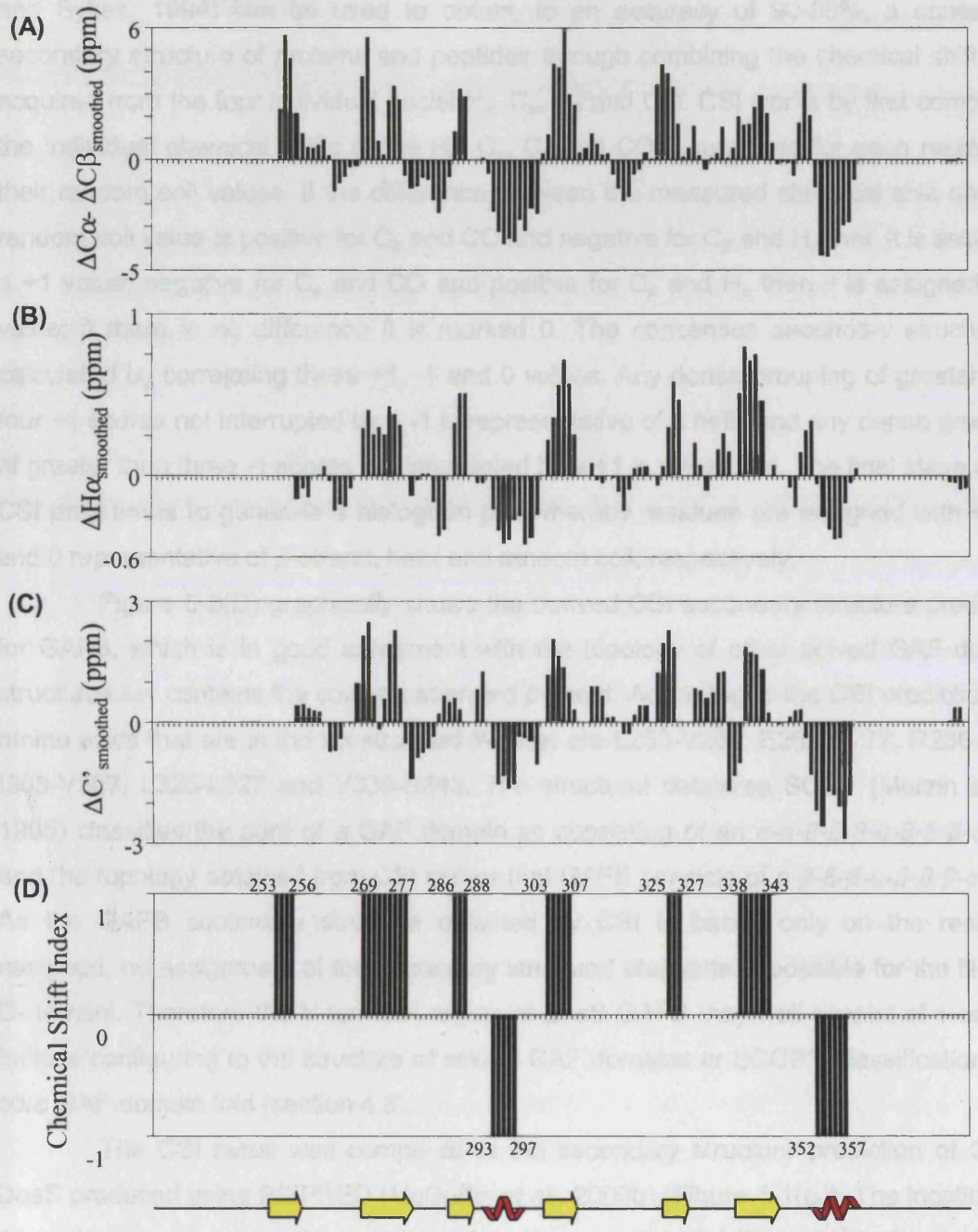


Figure 5-9: The chemical shift index for the prediction of the secondary structure of GAFB

(A) Secondary chemical shifts for the random coil values minus C_α-C_β values, ΔC_α-C_β

(B) Secondary chemical shifts for the H_α values minus random coil values, ΔH_α

(C) Secondary chemical shifts for the random coil values minus CO values, ΔCO

(D) CSI (Wishart et al., 1992; Wishart and Sykes, 1994) predicted secondary structure for each residue based on the chemical shift of the C_α, C_β, CO and H_α. Values >0 represent β-sheet, values <0 represent α-helix. The overall topology of GAFB is represented underneath the figure.

Blanks indicate the absence of a cross-peak or a residue that was too degenerate to determine assignment.

The software package Chemical Shift Index (CSI) (Wishart *et al.*, 1992; Wishart and Sykes, 1994) can be used to obtain, to an accuracy of 90-95%, a consensus secondary structure of proteins and peptides through combining the chemical shift data acquired from the four individual nuclei H_α, C_α, C_β and CO. CSI works by first comparing the individual chemical shifts of the H_α, C_α, C_β and CO resonances for each residue to their random coil values. If the difference between the measured chemical shift and the random coli value is positive for C_α and CO and negative for C_β and H_α then it is assigned a +1 value; negative for C_α and CO and positive for C_β and H_α then it is assigned a -1 value; if there is no difference it is marked 0. The consensus secondary structure is calculated by correlating these +1, -1 and 0 values. Any dense grouping of greater than four +1 scores not interrupted by a -1 is representative of a helix and any dense grouping of greater than three -1 scores not interrupted by a +1 is a β-strand. The final stage of the CSI program is to generate a histogram plot whereby residues are assigned with +1, -1 and 0 representative of β-strand, helix and random coil, respectively.

Figure 5-9(D) graphically shows the derived CSI secondary structure prediction for GAFB, which is in good agreement with the topology of other solved GAF domain structures *i.e.* contains the core six-stranded β-sheet. According to the CSI prediction the amino acids that are in the six-stranded β-sheet are L253-V256, E269-G277, R286-I288, I303-V307, L325-L327 and V338-S343. The structural database SCOP (Murzin *et al.*, 1995) classifies the core of a GAF domain as consisting of an α-α-β-β-β-α-β-β-β-α fold and the topology obtained from CSI shows that GAFB consists of a β-β-β-α-β-β-β-α fold. As the GAFB secondary structure obtained by CSI is based only on the residues assigned, no assignment of the secondary structural elements is possible for the N- and C- termini. Therefore the N-terminal region of DosS GAFB may well consist of 1 or 2 α-helices configuring to the structure of solved GAF domains or SCOP's classification of a core GAF domain fold (section 4.5).

The CSI result was compared to the secondary structure prediction of GAFB DosS produced using PSIPRED (McGuffin *et al.*, 2000b) (Figure 5-1(B)). The locations of the secondary structure elements based on the experimental ^1H and ^{13}C chemical shift data are, in general, well correlated with the α-helical and β-strand regions predicted by PSIPRED. PSIPRED did predict the N-terminal portion of GAFB to be α-helical, consistent with the structure of known GAF domains (Ho *et al.*, 2000; Martinez *et al.*, 2002). As GAFA from Phosphodiesterase 5 also contains two putative helices at the N- and C- termini (Sopory *et al.*, 2003), the hinge regions connecting the helices to the core β-sheet structure in both proteins may be flexible leading to the intermediate and slow conformational or chemical exchange seen in the [^1H,^{15}N]-HSQC spectra.

5.6 Overcoming the conformational/chemical exchange in GAFB

The [^{1}H,^{15}N]-HSQC spectrum of GAFB exhibits two major problems. First a number of the amide groups display multiple cross peaks with variable peak intensities suggesting the presence of multiple conformations, exchanging on a slow timescale. Second, the absence of a full set of NH peaks indicates that a portion of GAFB is undergoing conformational or chemical change on an intermediate timescale. In an effort to overcome the conformational/chemical exchange a number of factors were varied that might influence the appearance of the spectra. The results of those experiments are described within this section.

5.6.1 Refolding of GAFB

Refolding of GAFB was attempted for two reasons. First to verify that the multiple and missing NH peaks were not due to the protein being initially partially unfolded from *E. coli* expression. When the CD data obtained for GAFB was analysed by the program CONTIN the estimated helical content came to only 50 % of that predicted by PSIPRED (sections 5.1.1), possibly due to the N- and C- termini regions of GAFB being partially unfolded. Second, to confirm that the different structural conformations observed are not attributed to GAFB binding an endogenous ligand from *E. coli*.

A 0.6 mM [^{1}H,^{15}N]-labelled GAFB sample was unfolded in 6 M Guanidine-HCl and loaded onto a gel filtration column to remove any potential ligand that may have been present. Fractions containing GAFB were combined and concentrated. The protein was refolded by slow dilution into the NMR buffer (20 mM sodium phosphate, 100 mM NaCl, pH 6.0), full details are described in section 2.6.8. No protein precipitation was observed. A [^{1}H,^{15}N]-HSQC experiment was acquired before (on a 0.6 mM sample) and after (on a 0.1 mM sample) the refolding process; a section of both spectra are shown in Figure 5-10. Although some of the peaks have shifted, presumably due to the differences in concentration between the two protein samples, the peak shapes and the presence of multiple cross peaks are still observed, for example V284 and E274. Therefore, the refolding of GAFB did not improve the [^{1}H,^{15}N]-HSQC spectrum indicating that neither an endogenous ligand is present nor that refolding can isolate a single conformational state.

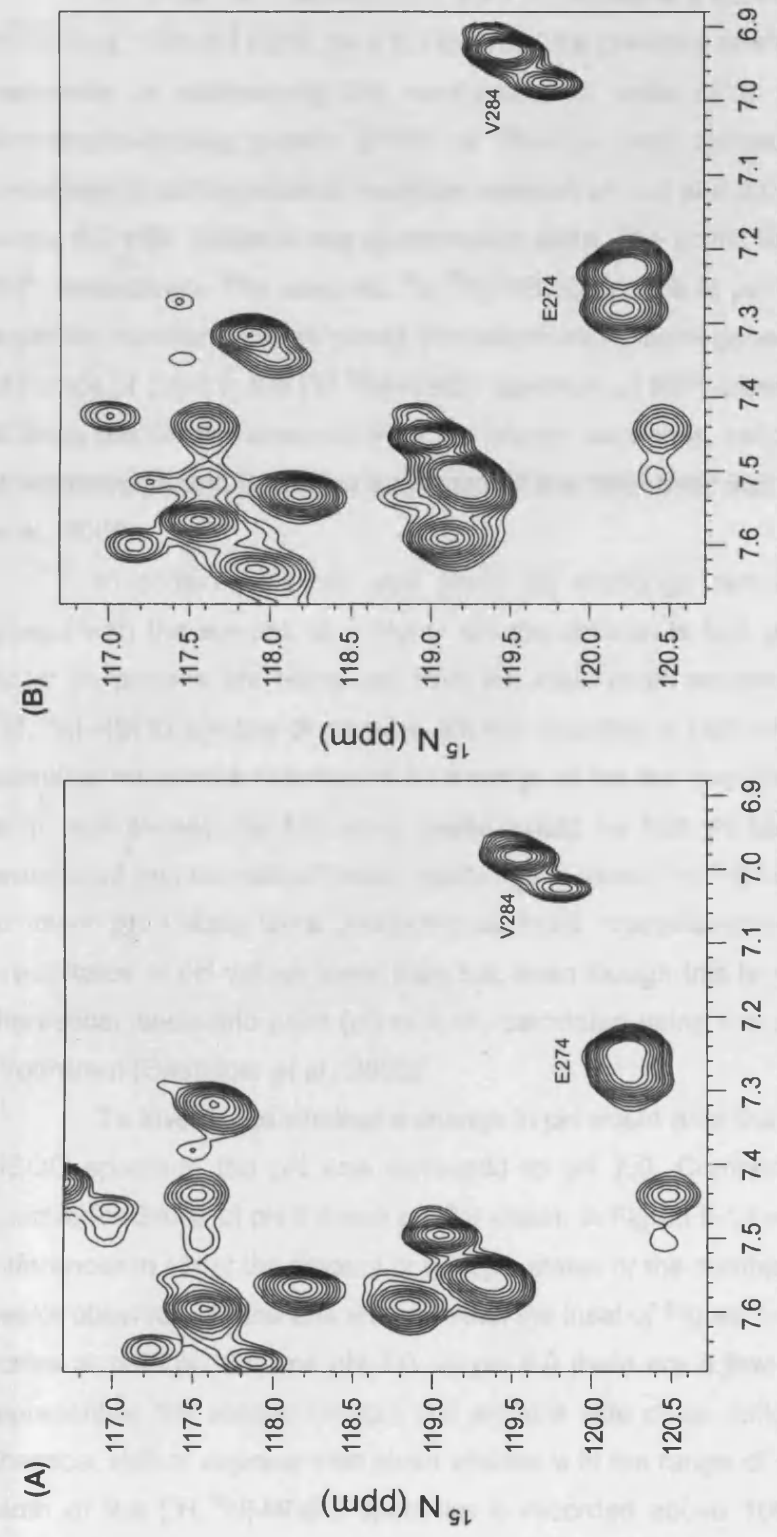


Figure 5-10: 2D Recorded ^{13}C-decoupled [^{1}H,^{15}N]-HSQC spectra of [^{13}C,^{15}N]-labelled GAFB in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 (A) before refolding (B) after refolding. All experiments were recorded on a 600 MHz Varian Inova spectrometer at 298K.

5.6.2 pH

All of the NMR experiments were conducted in a buffer comprising 20 mM sodium phosphate, 100 mM NaCl, pH 6.0. However, the pH-value of a solution can be an essential parameter in determining the conformational state of a protein. For example the pheromone-binding protein (PBP) of *Bombyx mori* (silkworm), as shown by NMR undergoes a conformational transition between pH 4.9 and 6.0. At a pH below 4.9 or a pH above 6.0 PBP exists in one conformation state, "the acidic form, (A)" or "the basic form, (B)", respectively. The acquired [^{1}H,^{15}N]-HSQC spectra at pH 4.5 and pH 6.5 contain the expected number of cross peaks consistent with a homogeneous sample. However, at a pH range of 5.0-6.0, the [^{1}H,^{15}N]-HSQC spectrum of PBP showed nearly twice the number of cross peaks than expected from the primary sequence, indicative of the protein existing in two forms, A and B in slow exchange on the NMR chemical shift time scale (Damberger *et al.*, 2000).

In addition, pH can also affect the exchange rate of protons from the amino groups with the solvent; at a higher pH the solution is less acidic and exchange rate is faster as protons are being lost from the main chain amides of proteins. Generally the [^{1}H,^{15}N]-HSQC spectra of proteins are not recorded at high pH values (> pH 8.0); as the chemical resonance recorded is an average of the two populations and the solvent water is in vast excess the NH cross peaks would be lost. At low pH values, residues are protonated and the rate of proton exchange is slow. [^{1}H,^{15}N]-HSQC experiments of GAFB at lower pH values were precluded as initial crystallisation trials revealed that GAFB precipitates at pH values lower than 5.8, even though this is ~1.5 pH units away from its theoretical isoelectric point (pI) of 4.45, calculated using the sequence analysis program ProtParam (Gasteiger *et al.*, 2005).

To investigate whether a change in pH would alter the appearance of the [^{1}H,^{15}N]-HSQC spectrum the pH was increased to pH 7.0. Comparison of the [^{1}H,^{15}N]-HSQC spectra for GAFB at pH 6.0 and pH 7.0 shown in Figure 5-11 reveals no overall significant differences in either the amount or multiple states or the number of cross peaks observed. Valine 284 shown within the inset of Figure 5-11 is present in at least two states at both pH 6.0 and pH 7.0. At pH 6.0 there are a few extra cross peaks possibly representing the aliased random coil arginine side chain amides (Rϵ) of GAFB. The ^{15}N chemical shift of arginine side chain amides is in the range of ~80 ppm and as the sweep width of the [^{1}H,^{15}N]-HQSC spectrum is recorded above 100 ppm their corresponding cross peaks are "wrapped around", *i.e.* aliased in the spectrum. Lower pH values (*i.e.* pH 6.0) reduces the exchange rate of solvent exposed amide protons in proteins with

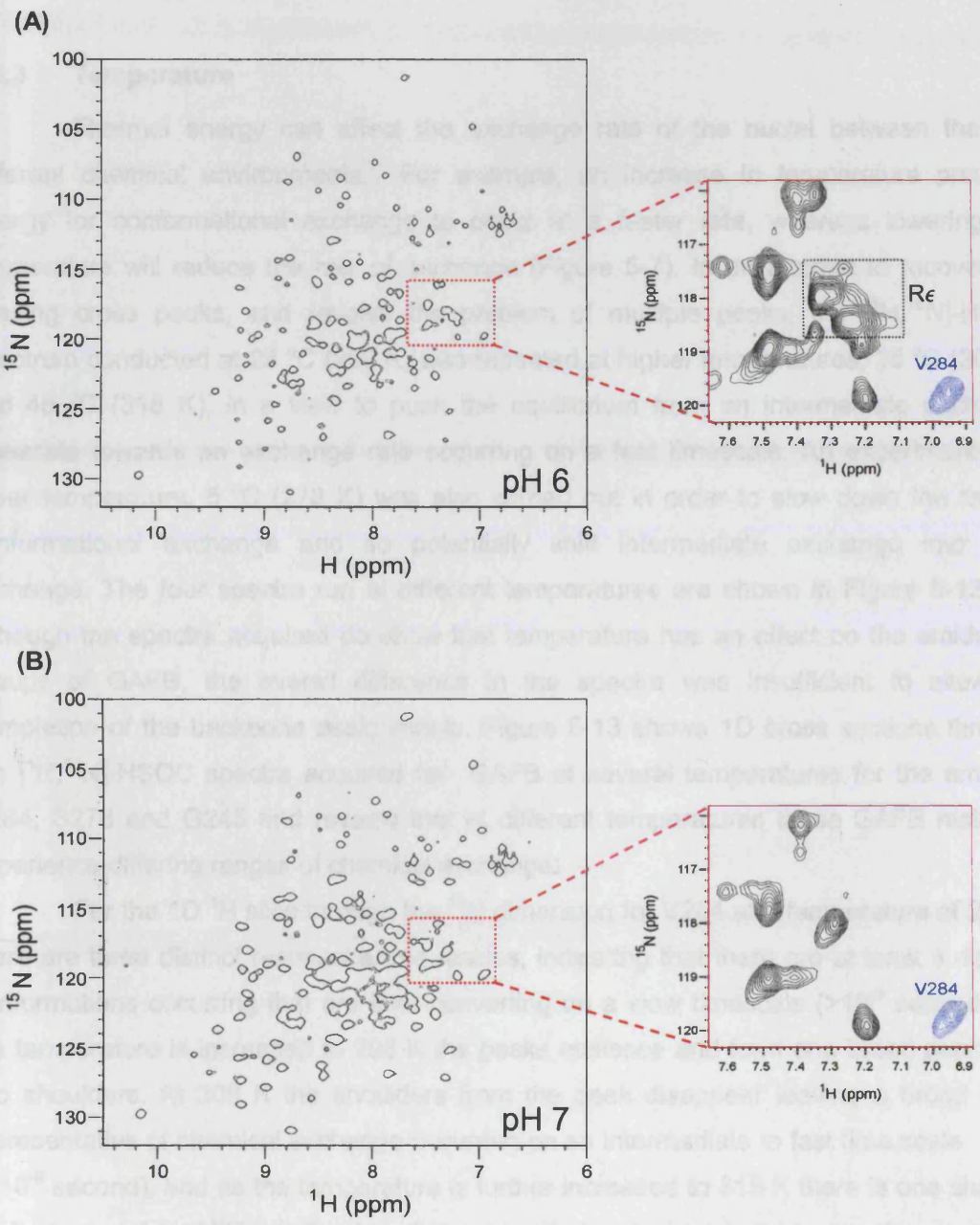


Figure 5-11: 2D [^{1}H,^{15}N]-HSQC spectra of [^{15}N]-labelled GAFB in 20mM sodium phosphate, 100mM NaCl, recorded on a 600 MHz Varian Inova spectrometer at 298K and different pH values. The aliased arginine random coil side amides (Rε) are labelled.

water protons, which results in solvent exposed amides, such as arginine side chains, to be observed in a [^{1}H,^{15}N]-HSQC spectrum.

5.6.3 Temperature

Thermal energy can affect the exchange rate of the nuclei between the two different chemical environments. For example, an increase in temperature provides energy for conformational exchange to occur at a faster rate, whereas lowering the temperature will reduce the rate of exchange (Figure 5-7). In an attempt to recover the missing cross peaks, and resolve the problem of multiple peaks, the [^{1}H,^{15}N]-HSQC spectrum conducted at 25 °C (298 K) was repeated at higher temperatures, 35 °C (308 K) and 45 °C (318 K), in a view to push the equilibrium from an intermediate exchange timescale towards an exchange rate occurring on a fast timescale. An experiment at a lower temperature, 5 °C (278 K) was also carried out in order to slow down the rate of conformational exchange and so potentially shift intermediate exchange into slow exchange. The four spectra run at different temperatures are shown in Figure 5-12 and although the spectra acquired do show that temperature has an effect on the amide NH groups of GAFB, the overall difference in the spectra was insufficient to allow for completion of the backbone assignments. Figure 5-13 shows 1D cross sections through the [^{1}H,^{15}N]-HSQC spectra acquired for GAFB at several temperatures for the amides, V284, S278 and G245 and reveals that at different temperatures these GAFB residues experience differing ranges of chemical exchange:

For the 1D ^{1}H slice through the ^{15}N dimension for V284 at a temperature of 278 K there are three distinct resonance frequencies, indicating that there are at least 3 distinct conformations occurring that are inter-converting on a slow timescale (>10^{-3} second). As the temperature is increased to 298 K the peaks coalesce and form one broad peak with two shoulders. At 308 K the shoulders from the peak disappear leaving a broad peak representative of chemical exchange occurring on an intermediate to fast time scale (10^{-3} to 10^{-6} second), and as the temperature is further increased to 318 K there is one sharper peak observed for V284, indicative of the rate of chemical exchange occurring on a fast timescale (<10^{-6} second), Figure 5-13(A).

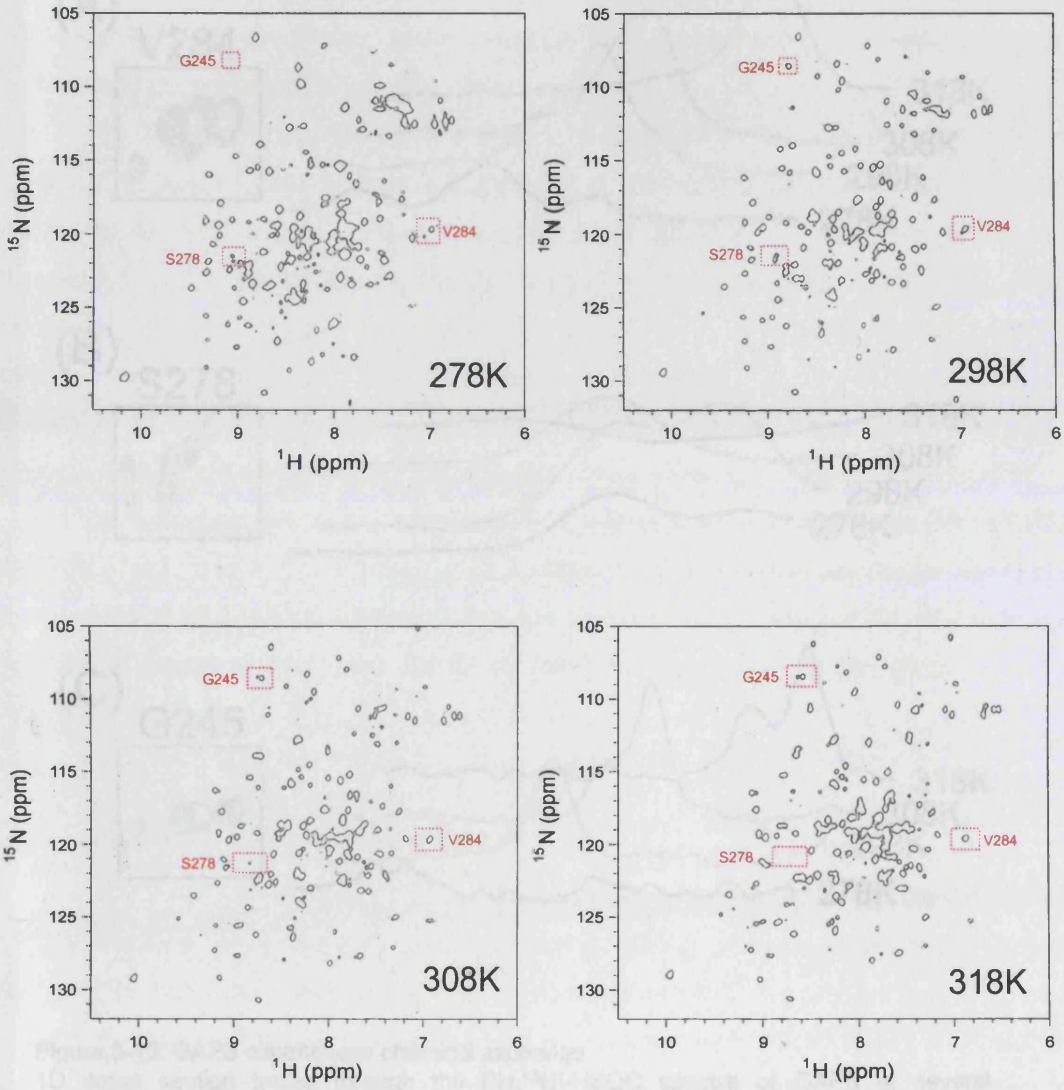


Figure 5-12: 2D ^{13}C-decoupled [^{1}H,^{15}N] HSQC spectra of [^{13}C,^{15}N]-labelled GAFB in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 recorded on a 600 MHz Varian Inova spectrometer at various temperatures.

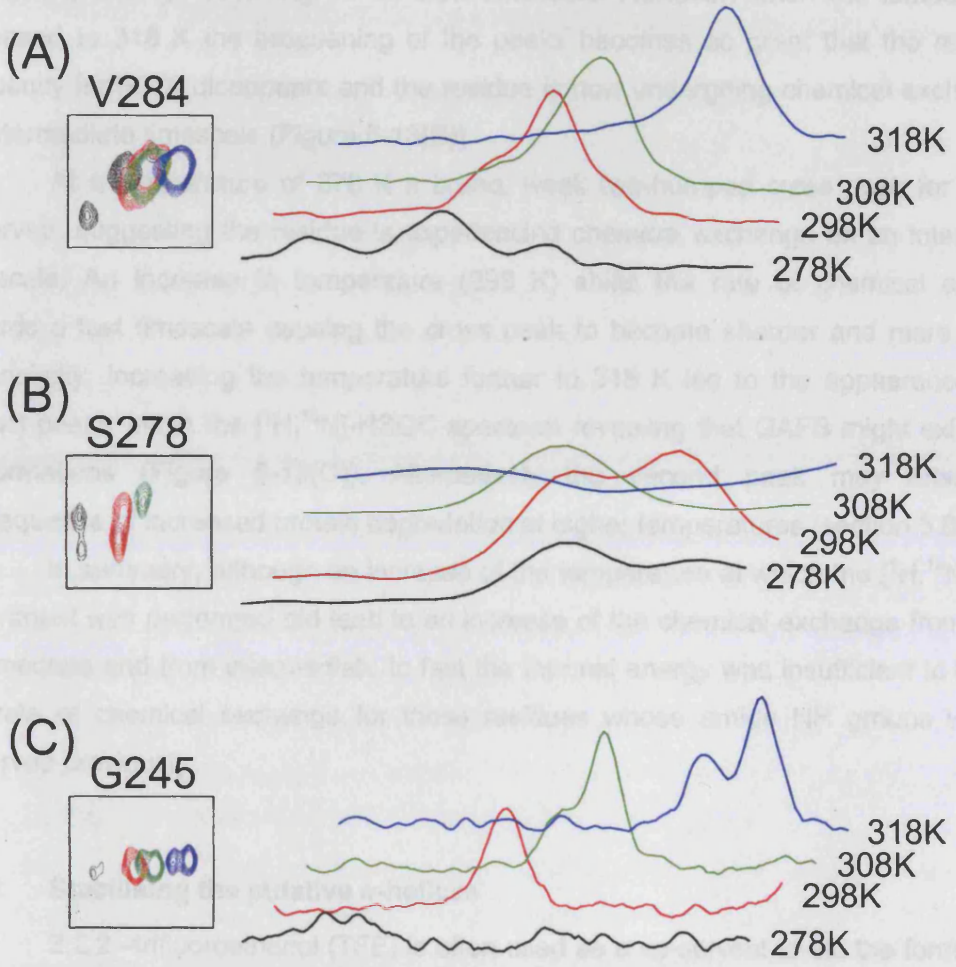


Figure 5-13: GAFB experiences chemical exchange
1D cross section traces through the [^{1}H,^{15}N]-HSQC spectra of GAFB at several temperatures for amides (a) V284, (b) S278 and (c) G245. The 1D slices for G245 and V284 are through the ^{1}H dimension, whilst for S278 they are taken in the ^{15}N dimension.
(A) V284 GAFB experiences differing ranges of chemical exchange from a slow timescale (>10^{-3} second) at 278K through to fast timescale (<10^{-6} second) above 308K.
(B) the amide S278 experiences slow exchange at temperatures below 298K and the broadening of the amide cross peak above 308K and disappearance at 318K is suggestive of chemical exchange on an intermediate timescale (10^{-6} to 10^{-3} second).
(C) for the amide cross peak of G245 the increase in temperature improves the appearance of the cross peak from one that is experiencing intermediate timescale at 278K to fast timescale at temperatures above 298K. However the increased resolution at temperatures above 308K reveals that the protein exists in either a second conformation or a second species due to degradation.

Initially at 278 K and 298 K the 1D ^{15}N slice through the ^{1}H dimension for S278 GAFB shows two broad peaks in the acquired [^{1}H,^{15}N]-HSQC spectra, representative of chemical exchange occurring on an slow timescale. However, when the temperature is increased to 318 K the broadening of the peaks becomes so great that the resonance frequency for S278 disappears and the residue is now undergoing chemical exchange on an intermediate timescale (Figure 5-13(B)).

At a temperature of 278 K a broad, weak two-humped cross peak for G245 is observed, suggesting the residue is experiencing chemical exchange on an intermediate timescale. An increase in temperature (298 K) shifts the rate of chemical exchange towards a fast timescale causing the cross peak to become sharper and more defined. Surprisingly, increasing the temperature further to 318 K led to the appearance of two distinct peaks within the [^{1}H,^{15}N]-HSQC spectrum revealing that GAFB might exist in two conformations (Figure 5-13(C)). Alternatively the second peak may arise as a consequence of increased protein degradation at higher temperatures (section 5.6.6).

In summary, although an increase of the temperature at which the [^{1}H,^{15}N]-HSQC experiment was performed did lead to an increase of the chemical exchange from slow to intermediate and from intermediate to fast the thermal energy was insufficient to increase the rate of chemical exchange for those residues whose amide NH groups were not observed previously.

5.6.4 Stabilising the putative α-helices

2,2,2 –trifluoroethanol (TFE) is often used as a co-solvent to aid the formation of, and to stabilise, secondary structure formation during peptide and protein structural studies. NMR and CD investigations have demonstrated that the presence of TFE assists and stabilises the formation of α-helix and β-sheet elements (Buck, 1998), TFE stabilises peptides by displacing water from the surface of peptides, which removes alternative hydrogen-bonding partners, thereby promoting the formation of intrapeptide hydrogen bonds (Roccatano *et al.*, 2002). TFE was added to a 0.6 mM sample of [^{1}H,^{15}N]-labelled GAFB to a final concentration of 30%, in the hope of trying to stabilise the predicted helical regions at the C- and N- termini. The addition of TFE caused no significant effects to the GAFB [^{1}H,^{15}N]-HSQC spectrum.

5.6.5 Ligand Screening

Ligand binding may cause large regions of a protein to undergo conformational exchanges. For example the guanine nucleotide-binding protein Ras occurs in two different conformational states when complexed with GTP analogues, cycling between a GDP-bound 'off' and a GTP-bound 'on' state (Spoerner *et al.,* 2005).

One of the unique aspects of NMR spectroscopy is that it provides a set of residue-specific probes with which to study protein-ligand interactions. The source for these probes is the ^{1}H-^{15}N correlation observed in the [^{1}H,^{15}N]-HSQC experiment (Clarkson and Campbell, 2003; Zuiderweg, 2002). The chemical shifts of the ^{15}N and ^{1}H atoms are sensitive to their local chemical environment and so a ligand interaction would cause the chemical environment to change around residues located at or near the binding interface. Therefore the chemical shifts for free and bound states are different and hence when a ligand is added to a protein, any binding event can be visualised in the [^{1}H,^{15}N]-HSQC spectrum as specific cross peaks shift position or broaden.

Identifying a ligand for GAFB may induce stabilisation of the regions of GAFB that are undergoing conformational exchange and thereby reduce the number of multiple states seen for one NH peak and recover the missing cross peaks.

Although GAF domains have been described as one of the largest families of small-molecule-binding regulatory domains, the direct evidence of ligand binding to GAF domains has only been established for a few ligands (Zoraghi *et al.,* 2004). Crystal structures of mouse phosphodiesterase 2A, cyanobacteria adenylyl cyclase cya B2 and a *D. radiodurans* photoreceptor DrCBD have been solved complexed with cGMP, cAMP, and bilin, respectively (section 4.5) and biochemical analysis have shown the binding of 2-oxoglutarate, pyruvate and formate (Zoraghi *et al.,* 2004). Although a motif search of the full length DosS sequence using the database PPsearch (European Bioinformatics Institute) (section 4.3) did not detect any of the ligand binding motifs contained within this database, ligand screening was attempted any way.

As cGMP has been shown to bind GAF domains with a high affinity, the Kd for the binding of a isolated GAFA domain from bovine PDE5 to cGMP has been reported to be 600 nM (Ho *et al.,* 2000) the ligands known to interact with GAF domains from other proteins were added to a solution of GAFB to a final concentration of 1 mM. Known GAF ligands: cGMP, cAMP, 2-oxoglutarate, pyruvate and formate were added to a 0.6 mM ^{15}N-labelled GAFB sample. An [^{1}H,^{15}N]-HSQC spectrum for each sample was recorded.

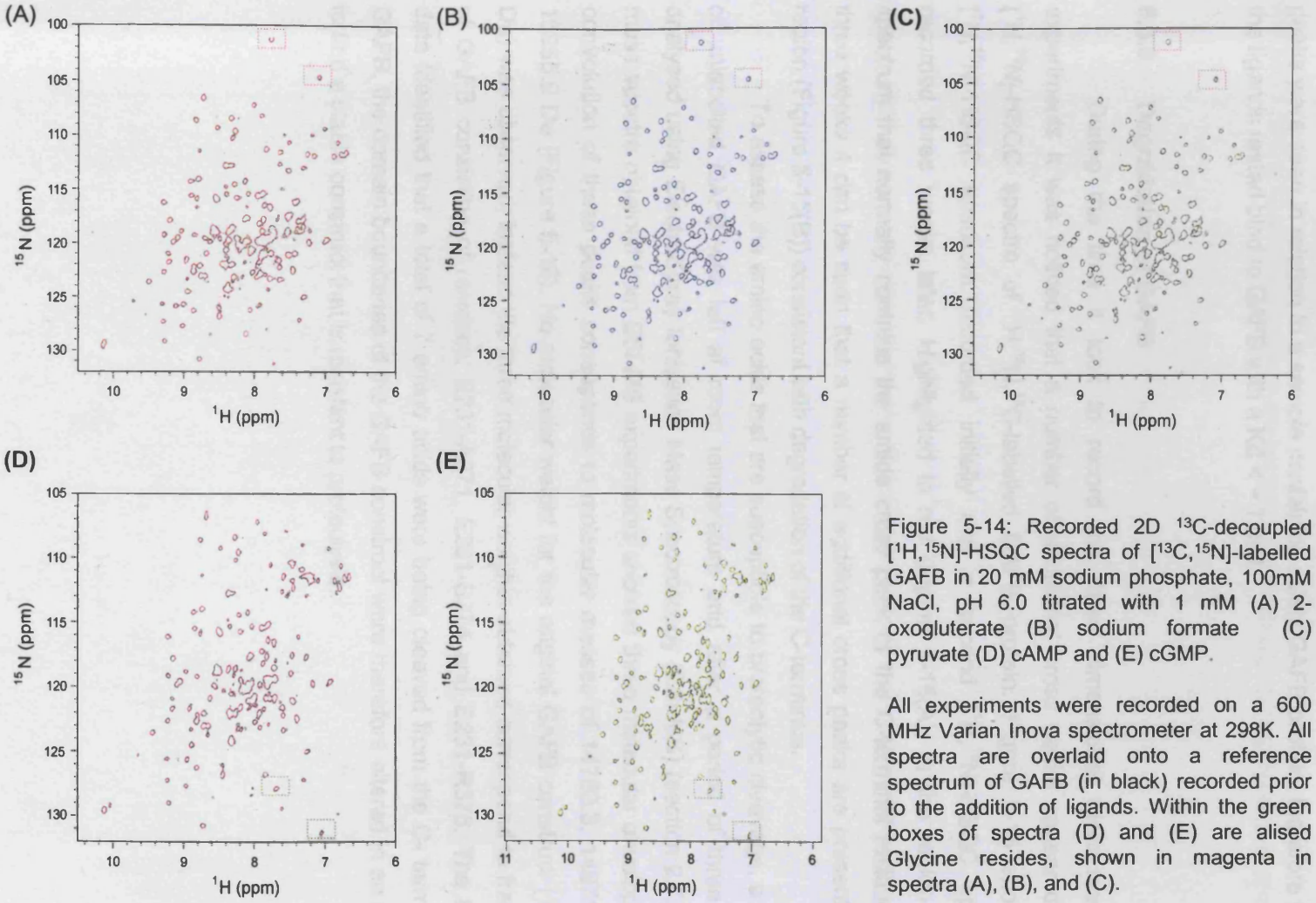


Figure 5-14: Recorded 2D ^{13}C-decoupled [^{1}H,^{15}N]-HSQC spectra of [^{13}C,^{15}N]-labelled GAFB in 20 mM sodium phosphate, 100mM NaCl, pH 6.0 titrated with 1 mM (A) 2-oxogluterate (B) sodium formate (C) pyruvate (D) cAMP and (E) cGMP.

All experiments were recorded on a 600 MHz Varian Inova spectrometer at 298K. All spectra are overlaid onto a reference spectrum of GAFB (in black) recorded prior to the addition of ligands. Within the green boxes of spectra (D) and (E) are alised Glycine resides, shown in magenta in spectra (A), (B), and (C).

The spectra are illustrated in Figure 5-14 and although slight differences between spectra can be observed due to degradation (section 5.6.6), no perturbations of the [^{1}H,^{15}N]-HSQC peaks were seen in relation to a sample containing solely GAFB protein, therefore none of the ligands tested bind to GAFB with a Kd < ~ 1 mM.

5.6.6 Degradation of GAFB

During the time it took to record the three dimensional triple resonance experiments it was noticed that a number of additional cross peaks appeared in the [^{1}H,^{15}N]-HSQC spectra of ^{1}H,^{15}N,^{13}C-labelled GAFB protein. Figure 5-15 shows the [^{1}H,^{15}N]-HSQC spectrum recorded initially and a second [^{1}H,^{15}N]-HSQC spectrum recorded three weeks later. Highlighted in red (Figure 5-15(A)) is the region of the spectrum that normally contains the amide cross peak of the C-terminal residue. After three weeks it can be seen that a number of additional cross peaks are present in this region (Figure 5-15(B)) consistent with degradation of the C-terminus.

To assess the amino acids that are susceptible to proteolytic cleavage, a sample of unlabelled GAFB was left at room temperature and after a period of three weeks analysed using ElectroSpray Ionisation Mass Spectroscopy (ESI-MS) (section 2.7.4). The mass spectra obtained from ESI-MS experiments showed three molecular envelopes, de-convolution of these peaks corresponds to molecular masses of 14783.3, 14971.4 and 15255.9 Da (Figure 5-16). No molecular weight for the original GAFB construct (15699.1 Da) was observed; instead the three molecular weights obtained correspond to fragments of GAFB consisting of residues: E231-A371, E231-S374 and E231-R376. The ESI-MS data identified that a total of 7 amino acids were being cleaved from the C- terminus of GAFB, the domain boundaries of the GAFB construct were therefore altered in an attempt to find a stable construct that is resistant to proteolysis.

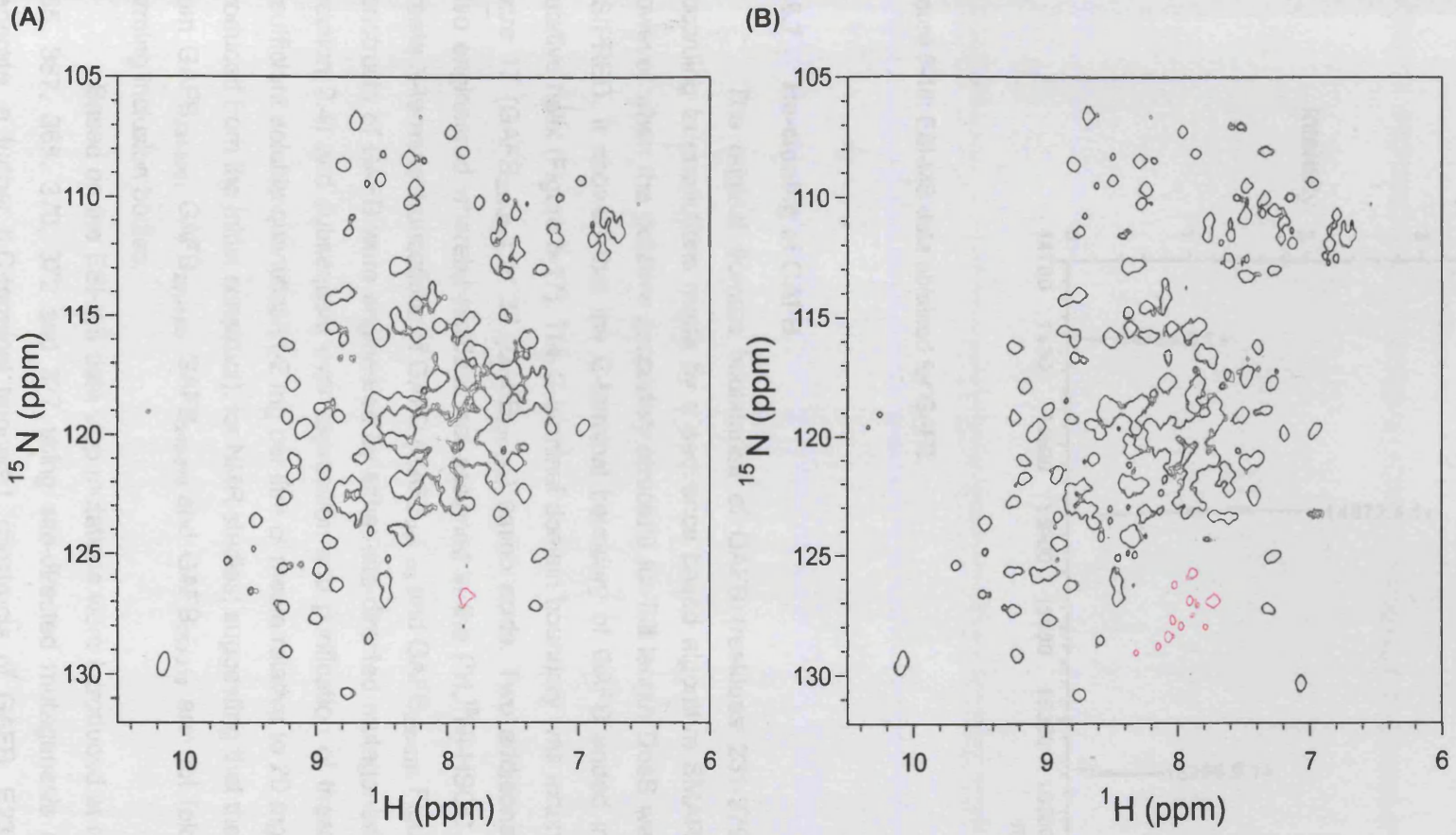


Figure 5-15: Recorded.2D ^{13}C-decoupled [^{1}H,^{15}N]-HSQC spectra of [^{13}C,^{15}N]-labelled GAFB in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 acquired before the start (A) and at the end off of (B) the triple resonance experiments. All experiments were recorded on a 600 MHz Varian Inova spectrometer at 298 K.

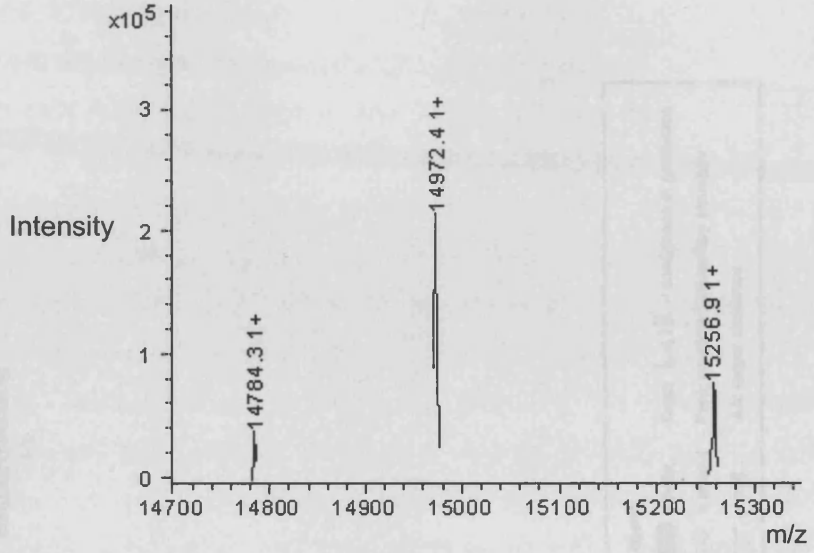


Figure 5-16: ESI-MS data obtained for GAFB.

5.6.7 Re-cloning of GAFB

The original domain boundaries of GAFB (residues 231-379) were designed according to predictions made by a sequence based algorithm SMART (section 3.1.2), however when the putative secondary structure for full length DosS was predicted using PSIPRED, it showed that the C-terminal boundary of GAFB ended in the middle of a putative helix (Figure 5-17). The C-terminal domain boundary was extended to include an extra 17 ($GAFB_{231-396}$) or 37 ($GAFB_{231-416}$) amino acids. Two additional constructs were also engineered whereby residues not assigned in the [^{1}H,^{15}N]-HSQC were removed to create N-terminal truncations of GAFB ($GAFB_{243-379}$ and $GAFB_{250-379}$, Figure 5-18). The four constructs of GAFB were engineered by either site-directed mutagenesis or PCR cloning (section 2.4) and subsequent over expression and purification of these proteins yielded insufficient soluble quantities (1-2 mg per litre of media relative to 20 mg per litre of media produced from the initial construct), for NMR studies; suggesting that the protein produced from $GAFB_{231-396}$, $GAFB_{231-416}$, $GAFB_{243-379}$ and $GAFB_{250-379}$ are not folding correctly and forming inclusion bodies.

Based on the ESI-MS data stop mutations were introduced at residues 361, 363, 365, 367, 368, 370, 372 and 377 using site-directed mutagenesis (section 2.4.8) to generate a further 8 C-terminal truncated constructs of GAFB, E231 to A360, A362,

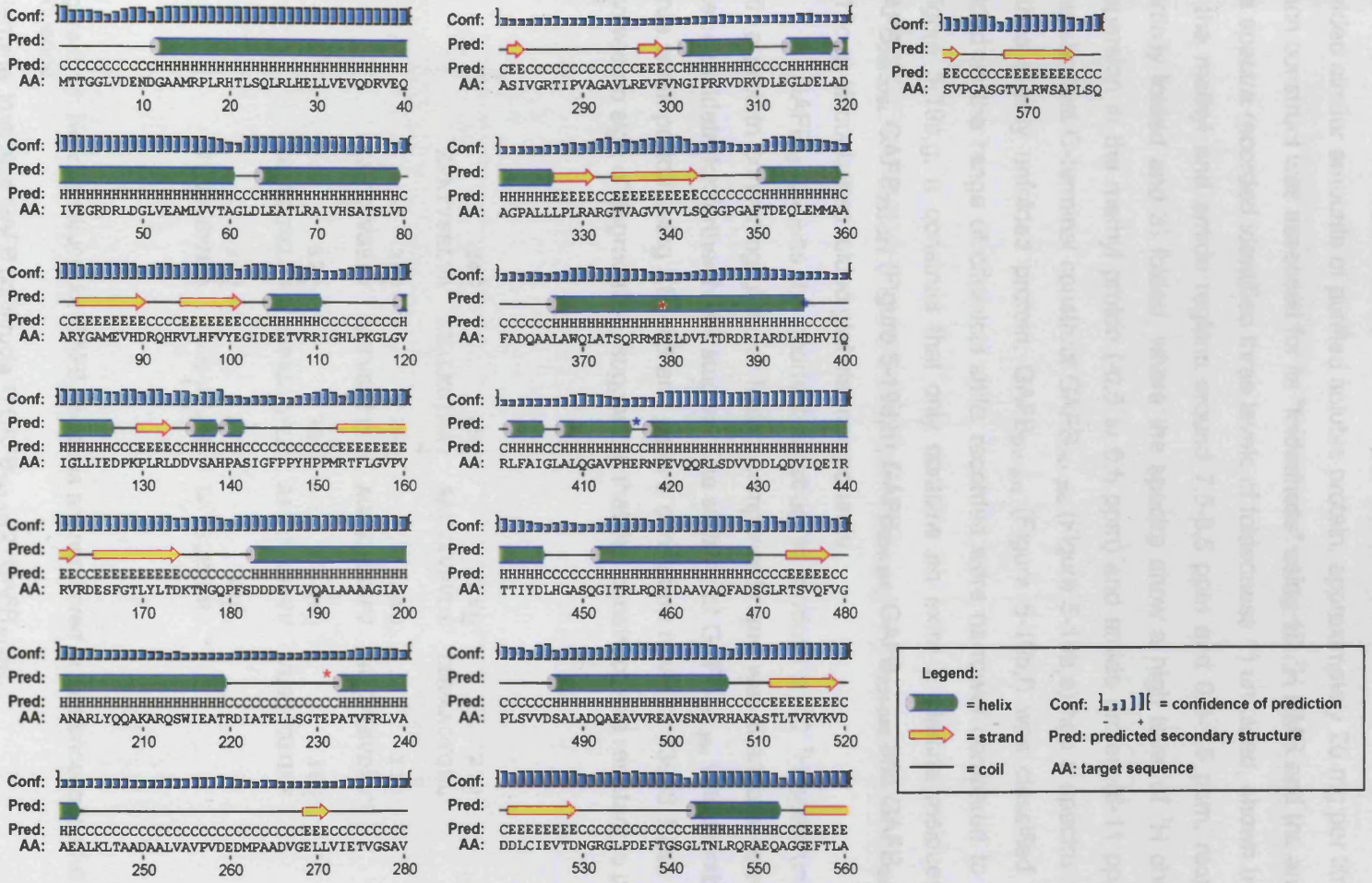


Figure 5-17: Putative secondary structure of full length DosS produced using PsiPred (McGuffin *et al.*, 2000), the original domain boundaries of $GAFB_{231-379}$ are marked with * , new C terminal truncations of GAFB are marked * at residues 396 and 416.

Q364, A366, L367, W369, L371 and R376, respectively, (Figure 5-18). The unlabeled proteins were expressed and purified in the same manner as GAFB (section 2.5.2), and yielded similar amounts of purified soluble protein, approximately 20 mg per litre of media. Each construct was assessed for its "foldedness" using 1D ^{1}H NMR and the analysis of all the spectra recorded identified three levels of foldedness 1) unfolded, shown by clustering of the methyl and amide regions around 7.5-8.5 ppm and 0.8-1.5 ppm, respectively 2) partially folded and 3) folded, where the spectra show a high level of ^{1}H chemical shift dispersion in the methyl proton (-0.5 to 0.5 ppm) and amide proton (6-11 ppm) regions. The shortest C-terminal construct $GAFB_{231-360}$ (Figure 5-19a,e) has a spectrum consistent with a mainly unfolded protein. $GAFB_{231-362}$ (Figure 5-19b,f) was classified as partially folded as the range of chemical shifts recorded were narrower compared to $GAFB_{231-364}$ (Figure 5-19c,g; a construct that only contains an extra 2 alanine residues). Proteins $GAFB_{231-376}$, $GAFB_{231-371}$ (Figure 5-19d,h), $GAFB_{231-369}$, $GAFB_{231-366}$ and $GAFB_{231-364}$ (Figure 5-19c,g) all contained folded globular components.

$GAFB_{231-364}$ was the shortest construct to produce a 1D ^{1}H spectrum consistent with a protein containing globular folded components and was therefore selected as the best candidate for further NMR studies. The stability of $GAFB_{231-364}$ was monitored over a three week period using ESI-MS and gave a consistent mass of 13958.1 Da. $GAFB_{231-364}$ showed no sign of degradation, suggesting that this construct was resistant to proteolysis.

```
        240         250         260         270
EPATVFRLVA  AEALKLTAAD  AALVAVPVDE  DMPAADVGEL
              1       2
        280         290         300         310
LVIETVGSAV  ASIVGRTIPV  AGAVLREVFV  NGIPRRVDRV

        320         330         340         350
DLEGLDELAD  AGPALLLPLR  ARGTVAGVVV  VLSQGGPGAF

        360         370         380
TDEQLEMMAA  FADQAALAWQ  LATSQRRMR
            3 4 5 67 8   9    10
```

Figure 5-18: New constructs for GAFB, residues are numbered in reference to the full length DosS sequence.

N-terminal truncated GAFB constructs 1) $GAFB_{243-379}$ 2) $GAFB_{250-379}$.

C-terminal truncated GAFB constructs, residues highlighted in red have been mutated to stop residues 3) $GAFB_{231-360}$ 4) $GAFB_{231-362}$ 5) $GAFB_{231-364}$ 6) $GAFB_{231-366}$ 7) $GAFB_{231-367}$ 8) $GAFB_{231-369}$ 9) $GAFB_{231-371}$ 10) $GAFB_{231-376}$.

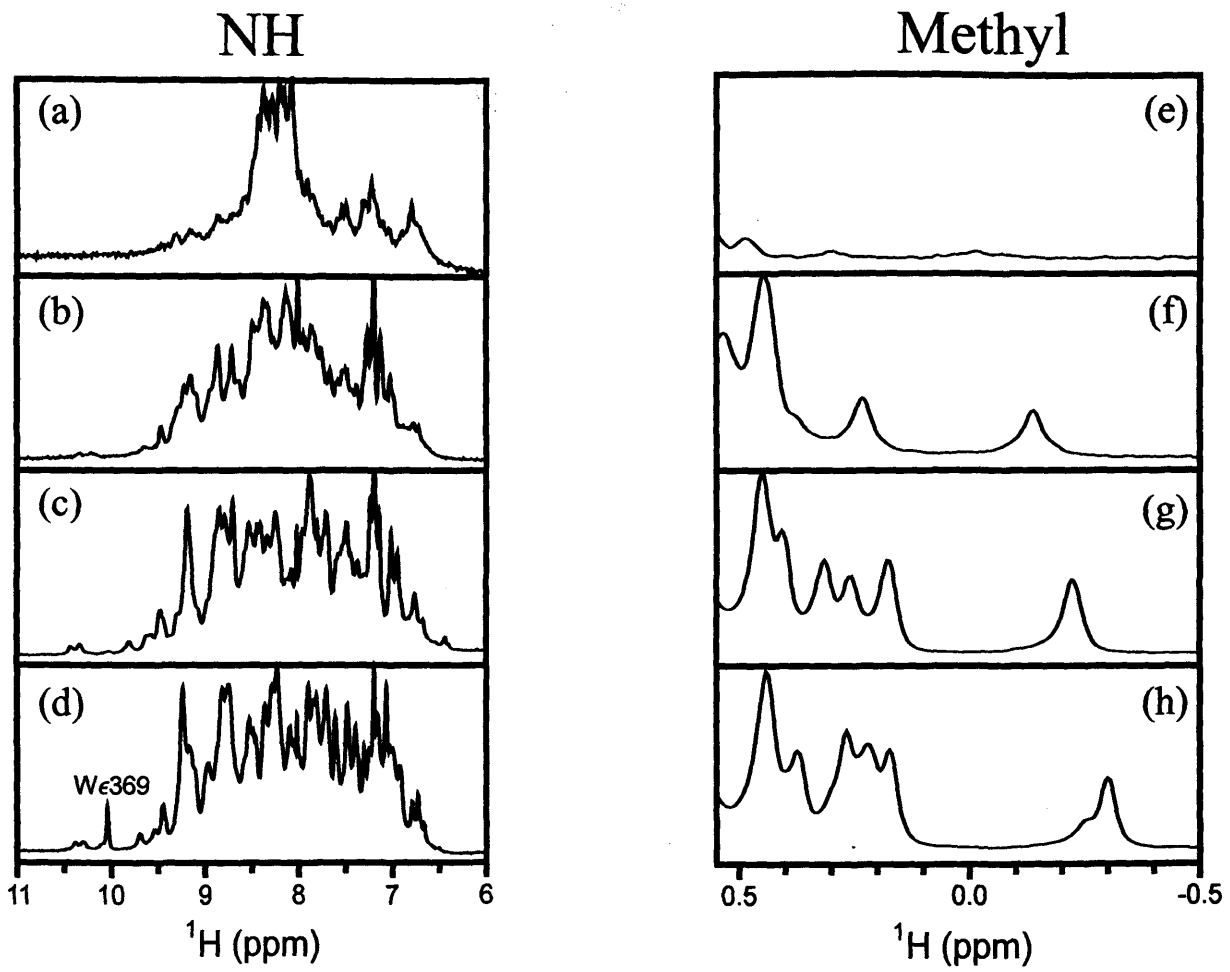


Figure 5-19: NH (a-d) and non-random coil methyl (e-h) regions of 1D ^{1}H NMR spectra for (a,e) $GAFB_{231-360}$, (b,f) $GAFB_{231-362}$, (c,g) $GAFB_{231-364}$ and (d,h) $GAFB_{231-371}$ recorded in 20 mM phosphate, 100 mM NaCl, pH 6.0.

5.6.8 [^{1}H,^{15}N] 2D Heteronuclear single quantum correlation spectroscopy (HSQC) of $GAFB_{231-364}$

^{15}N-labelled $GAFB_{231-364}$ protein was expressed in minimal media and purified in the same manner as ^{15}N-labelled $GAFB_{231-379}$ (section 2.5.3 and 2.6.1). A 0.6 mM sample of $GAFB_{231-364}$ containing 20 mM phosphate, pH 6.0, 10 % D_2O and 100 mM NaCl was used in an [^{1}H,^{15}N]-HSQC experiment performed at 25 °C. Analysis of the 2D [^{1}H,^{15}N]-HSQC spectrum of $GAFB_{231-364}$ (Figure 5-20) revealed no overall improvement in the quality of the backbone ^{15}N and ^{1}H chemical shifts resonances recorded. Although the cross peaks acquired for $GAFB_{231-364}$ are now of similar intensities to each other, from the inset shown in Figure 5-20, it is clearly visible that for V284 multiple cross peaks for a single amide are still present.

5.7 [^{1}H,^{13}C] 2D Heteronuclear single quantum correlation spectroscopy (HSQC) of $GAFB_{231-379}$

Essentially complete sequence specific backbone assignments of $GAFB_{231-379}$ were obtained for residues A250-A360, which represent a 110-residue structured core of the GAF domain in solution. Although assignments were not obtained for the N- and C-termini residues of GAFB, the possibility of determining the core β-sheet 3D structure of DosS GAFB was investigated further.

The assignment of side-chain NMR resonances is essential for solving a 3D protein structure. The amino acid sequence of GAFB contains a high proportion of methyl group-containing side chains residues A, I, L, T and V, which contribute to 53% of its amino acid sequence and only 2 aromatic residues within its core β-sheet. As aromatic residues contribute to dispersion of the methyl resonance cross peaks, and GAFB has such a low proportion of aromatic residues, it can be expected that an 2D-[^{1}H,^{13}C]-HSQC NMR spectrum of GAFB will show very little dispersion in the methyl region making the assignment of side chains problematic.

A ^{15}N,^{13}C-labelled $GAFB_{231-379}$ sample was used to record a 2D-[^{1}H,^{13}C]-HSQC NMR spectrum. The region of the [^{1}H,^{13}C]-HSQC spectrum corresponding to the methyl groups of GAFB is shown in Figure 5-21(A). The spectrum, as expected, shows limited resonance dispersion in both the ^{1}H and ^{13}C dimension. A substantial proportion of the cross peaks overlap in the central region of the spectrum (^{1}H:0.6-1.1 ppm;^{13}C:20-26 ppm). This overlap would cause great difficulties in assigning the methyl groups, and is a consequence of GAFBs amino acid sequence. When compared to [^{1}H,^{13}C]-HSQC spectra

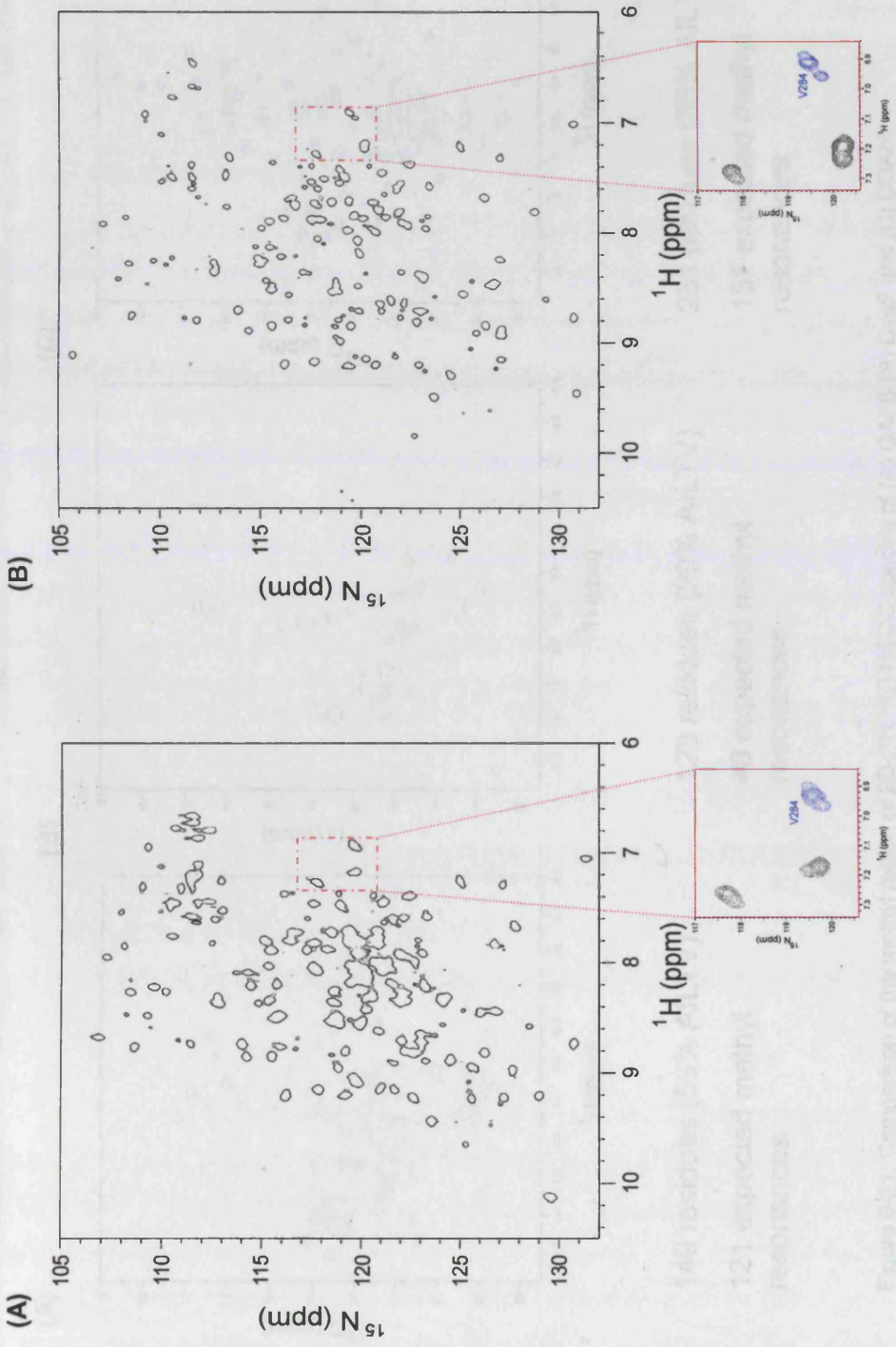


Figure 5-20: 2D [^{1}H,^{15}N]-HSQC spectrum of (A) [^{15}N]-labelled GAFB (B) [^{15}N]-labelled GAFB$_{231-364}$ in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 recorded on a 600 MHz Varian Inova spectrometer at 298 K.

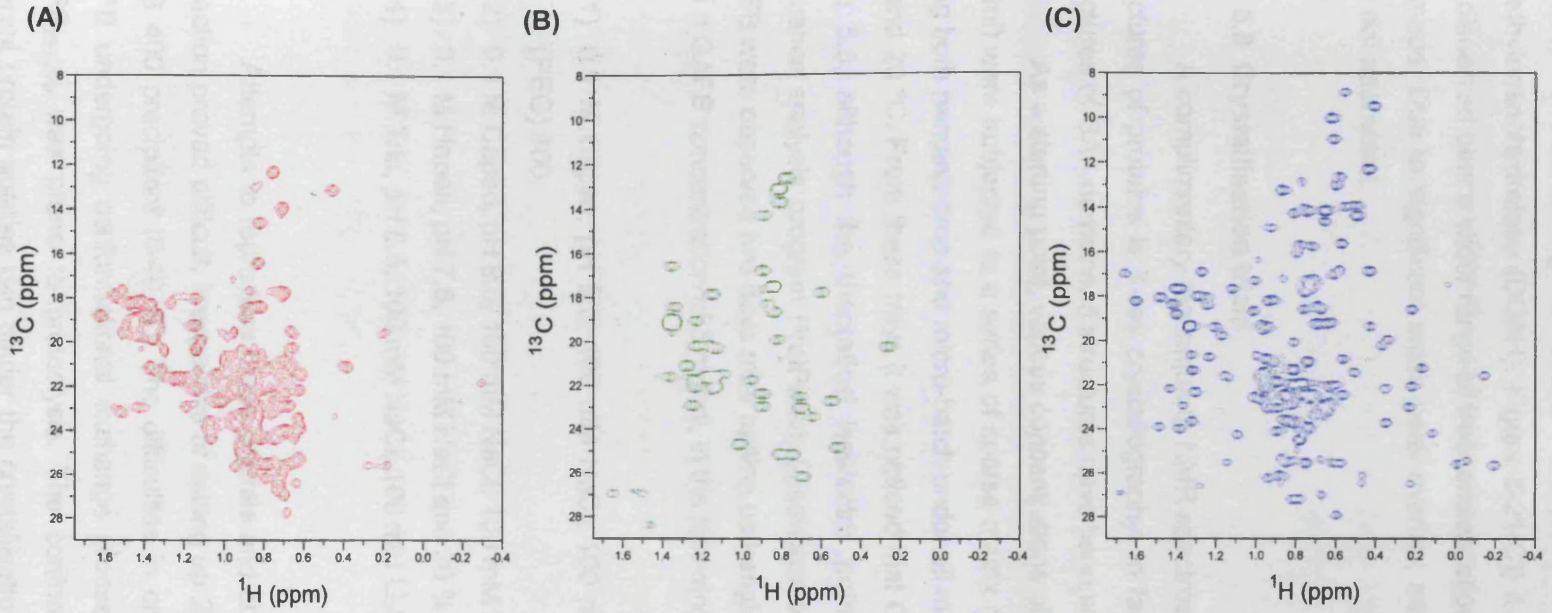


Figure 5-21: Comparison of the methyl region of 2D-[^{1}H,^{13}C] HSQC spectra of (A) GAFB (B) DysF and (C) DDAH.

of other protein domains of similar size (DysF domain from human myoferlin, Figure 5-21(B)) or containing a similar number of methyl groups (dimethylarginine dimethylaminohydrolase (DDAH), Figure 5-21(C)) it can be seen that the methyl groups are dispersed over a wider range of frequencies, allowing for unambiguous assignments to be made. Due to significant cross peak overlap, assignment of the sidechains of GAFB was not attempted.

5.8 Crystallisation trials

A complimentary technique to NMR spectroscopy in solving the three dimensional structures of proteins is X-ray crystallography. In fact, whilst there are no NMR solution structures of GAF domains, 5 structures have been solved by X-ray crystallography.

As a starting point, various concentrations of unlabelled DosS GAFB protein (2-20 mg/ml) were subjected to a series of sparse matrix crystallisation screens (section 2.7.5), using both hanging drop and micro-batch under oil methods. Trays were set up at 4 °C, 16 °C and 20 °C. From these trays it was noticed that GAFB precipitated at pH values lower than 5.8, although the theoretical isoelectric point (pI) is 4.45 calculated using the sequence analysis program ProtParam (Gasteiger *et al.*, 2005). Small needle crystals of GAFB were observed two days after setting up hanging drop trays, at all temperatures and with a GAFB concentration of 5 mg/ml, in the following conditions:

1) 0.1 M Capso, pH 9.5, 100 mM NaCl, 100 mM $MgCl_2$ and 30 % polyethylene glycol (PEG) 400.
2) 0.1 M Capso, pH 9.5, 100 mM NaCl, 100 mM Li_2SO_4 and 30 % PEG 400.
3) 0.1 M Hepes, pH 7.5, 100 mM NaCl and 30 % PEG 400.
4) 0.1 M Tris, pH 8.5, 100 mM NaCl, 100 mM Li_2SO_4 and 30 % PEG 400.

Attempts to reproduce these crystals and obtain larger crystals suitable for X-ray diffraction proved difficult, irrespective of setting up 2D grids varying pH (pH 6.5-10.5) and PEG 400 precipitant (5-40 %). The difficulties in crystal formation may be attributed to GAFB undergoing conformational exchange processes or the subsequent finding that $GAFB_{231-379}$ was prone to proteolysis. The conformational flexibility and presence of different protein species can hinder the crystallisation process. The tendency of dynamic proteins to crystallise may be significantly reduced as multiple conformers or protein

species exist in the crystallising solution. The conformer or protein species which produces a crystal suitable for X-ray diffraction may be present at significantly lower concentrations, the other conformations or species act as impurities perturbing crystallisation. Crystallisation attempts of GAF A domain from bovine PDE5 were initiated by Ho *et al.* but were not successful (Ho *et al.*, 2000). This finding may also be due to structural flexibility, as PDE5 GAFA has been found to exist in three different conformational states (Rybalkin *et al.*, 2003).

Although the shorter $GAFB_{231-364}$ construct showed resistance to proteolysis crystallisation trials were not attempted as the [^{1}H,^{15}N]-HSQC NMR spectrum acquired also revealed the presence of multiple cross peaks indicative of conformational exchange.

Currently, a thermofluor-based method is being developed within the lab to identify, for a particular protein, the optimal stabilising buffer conditions for crystallisation. The solvent buffer environment of a protein in a crystallisation experiment can be crucial to the formation of crystals, as it keeps the protein stable, minimises aggregation and therefore can control the events leading to crystallisation (Ericsson *et al.*, 2006). The idea behind this procedure is that folded and unfolded proteins can be distinguished through exposure to a hydrophobic fluoroprobe. The probe is quenched in aqueous solution but will preferentially bind the exposed hydrophobic interior of an unfolded protein leading to a sharp decrease in quenching so that a readily detected fluorescence emission can be investigated as a function of temperature. Melting temperatures of proteins obtained using the thermofluor methods have correlated well with values determined by other biophysical methods, for example CD (Ericsson *et al.*, 2006). In addition the thermofluor assay can be used to determine thermal melting curves of buffer/ligand conditions quickly in a high-throughput mode. Therefore, this method will be used to determine stable crystallisation buffers for DosS GAFB and other variants of GAFB and DosS, as well as being utilised to test a commercially bought fragment library of 500 low molecular weight compounds (Maybridge) for a putative ligand.

5.9 Summary

1D ^{1}H NMR and CD spectroscopic methods revealed that GAFB contained folded globular components. Efforts to determine the three dimensional structure of the GAFB domain from DosS was attempted through NMR spectroscopy and X-ray crystallography. These methods yielded no structure presumably due to the conformational flexibility of GAFB.

Analysis of the acquired [^{1}H,^{15}N]-HSQC spectrum for GAFB revealed the presence of multiple cross peaks with variable intensities and missing amide cross peaks. 78% of the backbone resonance assignments of GAFB were completed through the analysis of triple resonance NMR experiments. Due to incomplete data and signal overlap 22 of the observed cross peaks within the [^{1}H,^{15}N]-HSQC could not be assigned. In addition, the absence of 20 expected NH cross peaks in the [^{1}H,^{15}N]-HSQC spectrum could presumably be attributed to conformational exchange processes that occur on an intermediate timescale while the presence of multiple peaks might be due to chemical or conformational exchange occurring on a slow timescale. Examinations of the [^{1}H,^{15}N]-HSQC spectrum under a variety of sample conditions (temperature, pH, ligands, TFE) and attempts to re-clone GAFB failed to facilitate in the production of a fully assignable [^{1}H,^{15}N]-HSQC spectrum.

The putative structure of GAFB determined through CSI (using NMR data) and predicted using PSIPRED (using sequence homologues) reveals that the structure of GAFB contains a (α)-β-β-β-α-β-β-β-α fold that is consistent with previously solved GAF domain structures. The PSIPRED prediction for helical content of GAFB is 34% whereas from the analysis of CD data using the program CONTIN an estimate of 19 % helical content was determined. CONTIN is reported to be extremely accurate (95 %) for determining the percentage of helical content. Considering that the NMR data shows the N- and C- termini regions of GAFB undergoing conformational exchange and if these regions, predicted by PSIPRED to be helical but not assigned in the [^{1}H,^{15}N]-HSQC spectrum of GAFB are removed from the PSIPRED prediction the percentage of helical content of GAFB reduces to 13%, slightly lower than the CD estimation, thus providing a better agreement with the CD data-based estimation. This could suggest that, whilst the removal of the N- terminal α-helix produced an insoluble construct and so is essential for the overall stability of GAFB, the N-terminal predicted helix may not be stable within the isolated GAFB domain.

CHAPTER SIX

GENERAL CONCLUSIONS AND DISCUSSION

6 General conclusions and discussion

The initial aim of this thesis was to examine the structural properties of the input domains of three *MTB* sensory proteins (DosS, TrcS and MtrB), in order to identify their signalling function and therefore their roles within *MTB*. During the course of this thesis approximately 30 DNA constructs were generated and tested for expression; these constructs included full length TrcS, MtrB and DosS.

6.1 Domain Organisation

The cellular organisation of sensory input domains can be divided into three groups based on the presence and location of transmembrane regions (periplasmic, cytoplasmic and membrane-spanning), as discussed in the introduction (section 1.6.4). Analysis of the *MTB* MtrB and TrcS amino acid sequences using various algorithms (Chapter 3) generated the prediction of one and two transmembrane helices, respectively, within the N-terminal input domain; and a HAMP domain in the middle portion of each of the sequences. This prediction and the results from the purification of these two sensory MTB proteins (sections 3.1.1 and 3.3.1) suggests that both proteins are either periplasmic or membrane-spanning sensory proteins, with the HAMP domain providing a link to the cytoplasmic histidine kinase domain. The environmental signal(s) to which TrcS and MtrB respond are unknown and as bioinformatic analysis did not reveal the presence of any known putative domains and insufficient levels of stable soluble proteins of the full length constructs were obtained for structural studies, no further insight into the sensing function of these two proteins was revealed. In contrast, the expression of full length DosS and DosS variants within *E. coli* BL21 (DE3) pLysS cells resulted in the sufficient yield of solubly expressed proteins for further purification and characterisation (Chapter 2).

From the analysis of *MTB* DosS amino acid sequence (Chapter 3) two GAF domains (denoted GAFA: residues 63-210 and GAFB: residues 231-379) are predicted within the input sensing region. No transmembrane regions were predicted using TM-HMM even though previous studies identified three hypothetical transmembrane helices using the algorithm TM-Pred; one transmembrane helix was predicted between residues 187-206, within the GAFA domain, and two other helices, between residues 276-295 and 327-351, both of which are within the GAFB domain (Dasgupta *et al.*, 2000; Saini *et al.*, 2002). Expression of the DosS constructs resulted in the production of protein in both the insoluble and soluble fractions. However the purified soluble DosS proteins showed no

signs of aggregation even at high concentrations. In fact, Ioanaviciu and colleagues co-expressed GAFA and full length DosS with chaperones (GroEL/ES) and the expression of these proteins resulted solely in the soluble fraction, with higher yields of the DosS proteins obtained (16 mg/l of GAFA compared to 1.5 mg/l) (Ioanoviciu *et al.*, 2007). The NMR data on the isolated GAFB domain and secondary structure predictions for the full length DosS protein, also, do not indicate the presence of transmembrane helices. Namely the residues that were suggested to form transmembrane regions fall within the areas of observed or predicted β-strands and α-helices. Although the majority of sensory proteins that have been studied by others contain periplasmic sensing domains that are connected to their C-terminal cytoplasmic transmitter domains via a HAMP domain, the NMR data and PSIPRED prediction suggests that *MTB* DosS is not a membrane-spanning or periplasmic protein but a cytoplasmic, possibly membrane associated protein, sensing its stimulant from the cytoplasmic side of the membrane.

6.2 DosS binding haem

The identification of a high-spin b haem cofactor binding to histidine 149 in the GAFA domain of *MTB* DosS (Chapter 4) was not known prior to this work and was surprising, as no other GAF domain has previously been reported to bind haem. PAS domains, which commonly bind haem, are structurally similar to GAF domains, where the topology of both types of domain contains a core anti-parallel β-sheet surrounded by a varying number of α-helices. The main structural differences between these two domains occur in the location of a helix α^*/α^{**} and the number of β-strands in the core β-sheet, PAS domains have 5 β-strands and GAF domains have 6 β-strands. Although the secondary structure prediction program PSIPRED suggests the presence of only 5 β-strands within DosS GAFA, sequence analysis identifies the presence of a GAF domain within the same amino acid sequence. The missing β-strand in the PSIPRED prediction corresponds to β3, which in the structure of other GAF is the most varied β-strand within the core anti-parallel β-sheet varying in both position and length. However the PSIPRED predicted structural topology of GAFA consists of an α-β-β-α-α-β-α-β-β-α fold, which differs from the conventional topology, as defined by SCOP, of both a GAF domain (α-α-β-β-β-α-β-β-β-α) and a PAS domain (α-β-β-α-α-β-β-β). In fact the prediction appears to be a hybrid of both the PAS and GAF domain with the first part consistent with a PAS domain (α-β-β-α-α) and the second fragment a GAF domain (β-(α)-β-β-α).

The presence of a haem cofactor has implications for the understanding of the mechanism by which DosS senses external stimuli. DosS is known to be activated by

exposure to reduced oxygen tension and NO. The results presented in this thesis suggest that the regulation of DosS involves the binding of diatomic gases, such as O_2 and NO through the haem co-factor. As diatomic gases are capable of diffusing through the membrane the idea that DosS senses its stimulant from the cytoplasmic side of the membrane is plausible.

During the writing of this thesis the binding of diatomic gases to DosS has been published by Ioanoviciu and co-workers. Characterisation of full length DosS, GAFA and GAFA H149A mutant by UV-VIS and resonance Raman (RR) spectroscopy, confirmed that H149 coordinates the iron atom of the haem cofactor (Ioanoviciu *et al.*, 2007). Their work verifies that GAFA is able to reversibly bind O_2, CO and NO to the ferrous haem state and revealed that the H149A mutant was able to weakly bind haem but only when the growth media was supplemented with a haem substitute hemin, after induction. RR data showed that the H149A mutant had an altered haem environment that was unable to form a stable oxy complex and the formation of ferric-NO was also unstable. Characterisation of the distal and proximal haem binding pockets within the context of either the GAFA domain or full-length DosS was performed using RR spectroscopy. The ferrous and ferric states revealed no differences between the two proteins, suggesting that the proximal pocket (the imidazole ring of H149 binding to the haem molecule) in the full-length protein is not affected by interactions of GAFA with the other domains within DosS. However, differences were revealed in the distal binding pocket when CO was bound to the ferrous state. The RR data obtained for GAFA indicates that the H149 residue coordinates *trans* to the CO group and that two distinct protein species varying in distal pocket conformations are produced. However, the RR results for full-length DosS indicate the presence of only one of these CO-ferrous conformers. It was suggested that interdomain interactions within DosS cause a reduction in the flexibility around the haem co-factor and limits the number of conformations that could arise (Ioanoviciu *et al.*, 2007).

Collaborative efforts to investigate the impact of an in vivo H149A DosS mutant through complementation studies are underway. The DNA sequence of full length DosS, DosS H149A and DosT (a sensor that is paralogous to DosS) have been cloned into a pET30a vector with 1 kb DNA flanking regions on either side of the genes (section 2.4.4). These plasmids are being used by our collaborators to construct an isogenic mutant strain of *MTB* H37Rv using allelic replacement by homologous recombination. Allelic replacement involves the electroporation of nonreplicating DNA (suicide plasmids) carrying a mutated copy of the gene into the chromosome of *MTB*. A single recombination event between two alleles will result in integration of the whole plasmid to generate a strain carrying both wild-type and mutated copies of the gene. A second recombination event will

allow for the wild-type allele to be replaced by the mutant allele. The mutant *MTB* strain created will have the *dosT* gene disrupted (as it is paralogus to the *dosS* gene) by an antibiotic resistance determinant and the *dosS* gene replaced with a dosS H149A mutant. The effects of the mutation upon exposure to NO and hypoxia will then be investigated using real-time PCR and the data compared to the gene expression profiles already obtained for the wild-type strain by Stoker and colleagues (Kendall *et al.*, 2004). The mutant will then be complemented back to the original strain to confirm that any differences observed are due to the mutations created. Hopefully these complementation experiments will show that a single H149A mutation is capable of abolishing the role of DosS and will also validate the functional relevance of DosS *in vivo*.

6.3 Mechanism for DosS activation

The DosS/R signal transduction mechanism could be modelled using the sensory FixL and regulatory FixJ proteins. These two proteins are part of a two component regulatory system that control the expression of the nitrogen fixation genes in rhizobia and other bacteria, enabling the organisms to survive oxygen starvation. The FixL proteins are multi-domain cytoplasmic proteins. Although the number of domains present in the sensory region of the FixL proteins vary, located at the N-terminus is always a haem binding PAS domain. The input domain of *Bradyrhizobium japonicum* FixL protein is composed of two PAS and two PAC (motif C-terminal to PAS folds) domains as represented in Figure 6-1. Although haem binding occurs on His200 located in the first PAS domain, the importance of the other domains, for oxygen sensing and signal transduction is not known.

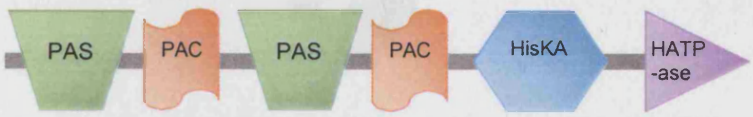


Figure 6-1: Schematic representation of domain organisation within the protein FixL from *Bradyrhizobium japonicum*.

Activity of the histidine kinase domain of FixL is inhibited strongly by binding of oxygen to an associated haem-containing PAS domain. In the absence of oxygen, FixL is autophosphorylated at an invariant histidine residue within the HisKA domain, the

phosphoryl group is subsequently transferred to FixJ, leading to enhanced transcriptional activity. The mechanism of kinase regulation by oxygen is not fully understood. However, crystal structures of un-ligated and various ligand-bound forms of FixL have been solved and reveal that the PAS fold, upon the binding of oxygen, undergoes minor structural changes in the vicinity of the haem pocket (Figure 6-2). The conformational switch involves the movement of Arg220 (within the β3 strand), which upon oxygen binding is released from the salt bridge with the haem moiety and moves to form a hydrogen bond with the bound oxygen molecule, associated with a shift in the loop region (FG loop) connecting α3 to β3 (Dunham *et al.*, 2003). These shifts are thought to be recognised by the kinase domain but the effects of these structural changes on controlling the autophosphorylation ability of FixL are not as yet understood. In addition it was revealed that binding of nitric oxide to FixL proteins suppresses HK activity but with only two fold compared to an inhibition factor of over 100 fold exhibited by oxygen (Dunham *et al.*, 2003; Tuckerman *et al.*, 2002).

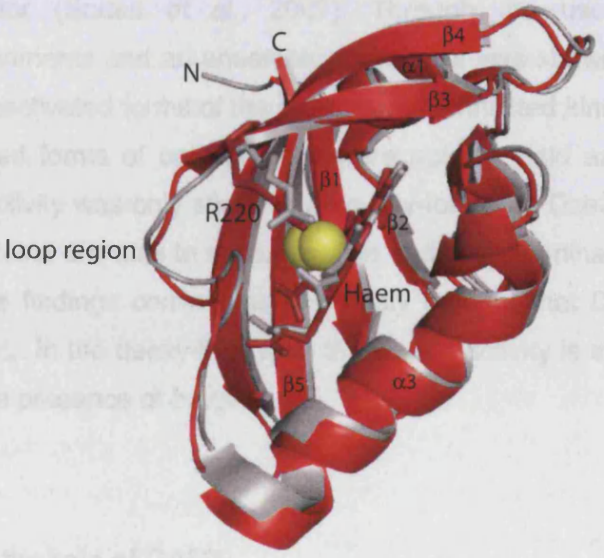


Figure 6-2: Binding of oxygen to FixL; in grey ligand free FixL (pdb 1DRM), in red FixL bound to a molecule of oxygen (pdb 1DP6), oxygen is coloured yellow. C-terminal helix α4 has been removed from both structures.

Analogously, the HK domain of DosS MTB may be inactivated by oxygen and activated under hypoxic conditions with associated induction of the DosRS regulon; the haem binding ligand nitric oxide would also be able to bind and activate DosS. Although

the ^{32}P-ATP auto-phosphorylation experiments performed on full length DosS with and without the presence of NO are preliminary, they would support this hypothesis. The DosS kinase domain, purified without resorting to denaturation for the first time, showed the ability to auto-phosphorylate and thus incorporate the ^{32}P labelled-γ phosphate group from ATP in the presence of Mg^{2+}. When the same experimental procedures were performed on full length DosS, the protein did not exhibit auto-phosphorylation activity and this ability of full length DosS to undergo auto-phosphorylation was restored only in the presence of diethylenetriamine-nitric oxide, an NO releasing compound (Chapter 4). As these experiments were unpredictable possibly because they were performed in an aerobic environment and the oxy-DosS form is thought to be the inactive form of the protein, plans to either purchase a hypoxic chamber or set up collaboration with another laboratory are underway. However, following the publication of DosS binding haem, the findings described in chapter 4, the laboratory of Marie-Alda Gilles-Gonzalez has investigated whether the ligands O_2, CO and NO could regulate autophosphorylation of DosT or DosS with the presence of ATP (Sousa *et al.*, 2007). Sousa and co-workers reported that both DosS and DosT are oxygen-switched kinases and they confirmed that DosT also functions via a haem co-factor (Sousa *et al.*, 2007). Through the use of ^{32}P-ATP auto-phosphorylation experiments and an anaerobic chamber it was shown that oxy-DosS and oxy-DosT were the inactivated forms of the proteins, with inhibited kinase activity, whereas CO, NO and unligated forms of both proteins were able to yield an active kinase. As inhibition of kinase activity was only shown for the oxy-forms for DosS and DosT the data reveals that both proteins are able to sense oxygen and to discriminate against regulation by CO or NO. These findings confirm, as previously thought that DosS functions in an analogous way to FixL. In the deoxy-FixL form the kinase activity is switched on, however it is switched off in the presence of oxygen.

6.3.1 Elucidating the role of GAFB

Nearly complete sequence specific backbone assignments of GAFB were obtained for residues A250-A360, except for the amino acids V290, V310-L312, E317, L329, R330 and G348.This portion of the sequence forms a β-β-β-α-β-β-β-α fold as derived by secondary chemical shift analysis and is consistent with the core topology of a GAF domain (α-β-β-β-α-β-β-β-α). Complete backbone assignments for all residues was not possible due to the presence of multiple cross peaks, missing cross peaks and ambiguous data (Chapter 5). The absence of NH cross peaks in the [^{1}H,^{15}N]-HSQC

spectrum is attributed to conformational or chemical exchange processes that occurs on an intermediate timescale and the presence of multiple peaks due to conformational exchange occurring on a slow timescale. As the majority of the unassigned cross peaks are part of the N- and C- terminal regions of GAFB, predicted to be α-helical, the conformational flexibility is associated within these regions. A number of possible explanations could be envisaged to explain the conformational heterogeneity observed in the acquired [^{1}H,^{15}N]-HSQC spectrum of GAFB and are summarised pictorially in Figure 6-3.

The first idea involves conformational flexibility about the hinge regions connecting the N- and C- terminal α-helices to the core β-sheet structure (Figure 6-3(A)). If the interactions between the helices and the core β-strand region are not energetically favourable then the flexibility in the hinge regions will cause the N- and C- terminal helices to sample conformational space. If the rate of exchange between the positions of the helices is occurring on an intermediate NMR time scale, then signal broadening would result with the NH cross peaks of residues in the helices disappearing from the [^{1}H,^{15}N]-HSQC spectrum. The movement of these helices towards and away from the core β-sheet would also create unique chemical environments for the NH groups within the β-sheet that form transient interactions with the helices. If the rate of exchange is slow on the NMR timescale, the [^{1}H,^{15}N]-HSQC spectrum will show distinct multiple populations of NH cross peaks resulting from the different chemical environments.

Analysis of the circular dichroism data acquired for GAFB provides an alternative explanation for the absence of expected cross peaks from the [^{1}H,^{15}N]-HSQC spectrum. The program CONTIN estimated the percentage of the α-helical content of GAFB to be lower than that predicted by PSIPRED (Chapter 5). If the N- and C- terminal α-helices of GAFB were disordered, or switching between folded and unfolded states at an intermediate rate of exchange, the cross peaks in the [^{1}H,^{15}N]-HSQC spectrum of GAFB, attributed to these regions, would be lost; and the predicted amount of α-helical content lowered, to nearer the estimated value. As the chemical environment seen by the β-strands is altered by whether the helices are folded or unfolded or transient interactions, multiple cross peaks would be observed.

Although the scenarios discussed above do provide explanations for the problems of absent or multiple cross peaks within the [^{1}H,^{15}N]-HSQC spectra of GAFB, there are two

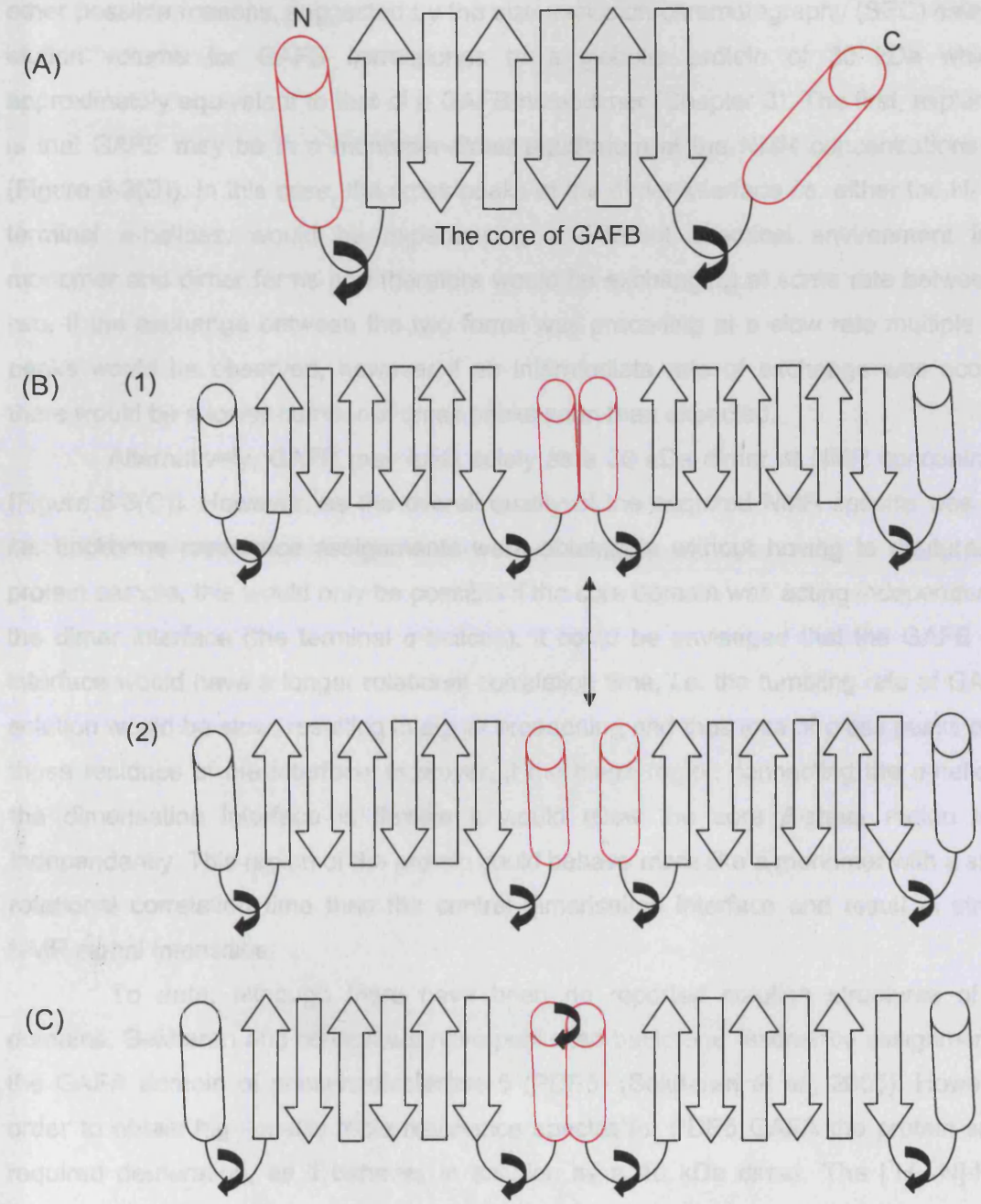


Figure 6-3: Explanations for the conformational heterogeneity seen in the [^{1}H,^{15}N]-HSQC spectra of DosS GAFB.

(A) Movement of the N- and C-terminal regions of GAFB, possibly creating conformation exchange.

(B) GAFB in a dimer (1) – monomer (2) equilibrium.

(C) GAFB as a dimer.

other possible reasons, suggested by the size exclusion chromatography (SEC) data. The elution volume for GAFB corresponds to a globular protein of 30 kDa which is approximately equivalent to that of a GAFB homodimer (Chapter 3). The first, explanation is that GAFB may be in a monomer-dimer equilibrium at the NMR concentrations used (Figure 6-3(B)). In this case, the cross peaks at the dimer interface *i.e.* either the N- or C-terminal α-helices, would be experiencing a different chemical environment in the monomer and dimer forms and therefore would be exchanging at some rate between the two. If the exchange between the two forms was preceding at a slow rate multiple cross peaks would be observed, however if an intermediate rate of exchange was occurring there would be a lower number of cross peaks seen than expected.

Alternatively, GAFB may exist solely as a 30 kDa dimer at NMR concentrations (Figure 6-3(C)). However, as the overall quality of the acquired NMR spectra was good, *i.e.* backbone resonance assignments were obtainable without having to deuterate the protein sample, this would only be possible if the core domain was acting independently of the dimer interface (the terminal α-helices). It could be envisaged that the GAFB dimer interface would have a longer rotational correlation time, *i.e.* the tumbling rate of GAFB in solution would be slow, resulting in signal broadening and thus loss of cross peaks of only those residues of the interface. However, if the hinge region connecting the α-helices at the dimerisation interface is flexible it would allow the core β-sheet region to act independently. This region of the protein could behave more like a monomer with a shorter rotational correlation time than the central dimerisation interface and result in stronger NMR signal intensities.

To date, although there have been no reported solution structures of GAF domains, Sekharan and co-workers have published backbone resonance assignments for the GAFA domain of phosphodiesterase 5 (PDE5) (Sekharan *et al.,* 2005). However in order to obtain high-quality triple resonance spectra for PDE5 GAFA the protein sample required deuteration, as it behaves in solution as a 46 kDa dimer. The [^{1}H,^{15}N]-HSQC spectrum acquired had 38 expected backbone amide resonances missing and the majority of the unobserved and/or unassigned cross peaks in the [^{1}H,^{15}N]-HSQC spectrum of PDE5 GAFA are located in the same region of the protein as in DosS GAFB, at the N- and C-termini.

The conformational flexibility of DosS GAFB and PDE5 GAFA may be an inherent feature of GAF domains, and important for its function. As previously mentioned, in the PAS domain of FixL, ligand binding to the haem iron changes the position of several residues at the haem distal site, including the loop region connecting helix α-3 to strand β-3 (the FG loop) (Figure 6-2). It is believed that these structural changes around the haem

region are a key factor in kinase regulation. Therefore, the binding of different diatomic gas to DosS GAFA may induce slight structural changes within the microenvironment of the haem distal binding pocket causing GAFB to undergo further structural perturbations and perhaps amplify these movements to hence generate a structural cascade and feed information to the histidine kinase domain (HisKA/ Hpt/DHp). The loop region connecting GAFA to GAFB may provide flexibility and allow for different orientations of GAFA in relation to GAFB, depending on the state of the haem iron, these structural changes might propagate through the sensory protein to regulate the activity of the HisKA domain, by possibly promoting dimerisation of the HisKA domain and allowing for trans-phosphorylation to occur (Figure 6-4). In two-component signal transduction systems the ATP-dependant phosphorylation of a conserved His residue within the HisKA domain can occur via a *trans*-autophosphorylation mechanism. One HisKA monomer phosphorylates a second monomer within the HisKA dimer. The phosphoryl group from HisKA is then transferred to an Asp residue in the cognate response regulator (Robinson *et al.*, 2000).

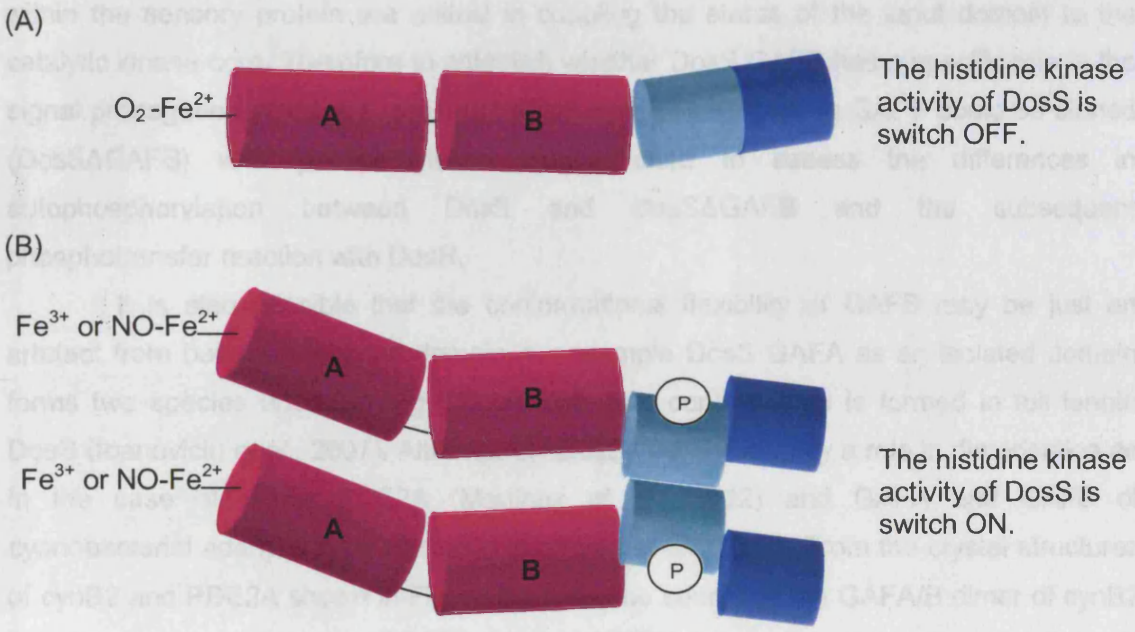


Figure 6-4: A possible mechanism for DosS regulation. The input domain of DosS contains two GAF domains coloured magenta, the oxy-DosS state is inactive (A), whereas the deoxy or NO bound DosS is active (B) capable of switching on histidine kinase activity through dimerisation of the HisKA domain, coloured light blue. The ATP binding domain (HATPase) is shown in dark blue.

Marina and co-workers determined the crystal structure of a sensory protein fragment consisting of the HisKA and HATPase from *T.maritima TM0853.* The structure was solved as a dimer, with the dimer interface associated exclusively between the antiparallel helices of the HisKA domain. As the ATP binding site within the HATPase domain is distant from the phosphorylated His residue in the HisKA helix they postulated that the HATPase domain of the sensor must access multiple conformational states in order for trans-autophosphorylation to occur (Marina *et al.,* 2005).

Within two-component systems HAMP domains have also been reported to undergo conformational changes in order to propagate the status of an input domain to its associated histidine kinase domain. HAMP domains link the periplasmic input domain to a cytoplasmic histindine kinase domain through the membrane. Inouye and co-workers describe the solution structure of a HAMP domain of the transmembrane receptor Af1503 from *Archaeoglobus fulgidus* and suggest that it transduces signals through the membrane via a simple rotation of its four-helix parallel coiled coil (Inouye, 2006).

Both of the examples described, support the idea that conformational changes within the sensory protein are critical in coupling the status of the input domain to the catalytic kinase core. Therefore to establish whether DosS GAFB has a specific role in the signal propagation process a construct of full length DosS lacking GAFB could be cloned (DosSΔGAFB) with phosphorylation assays used to assess the differences in autophosphorylation between DosS and DosSΔGAFB and the subsequent phosphotransfer reaction with DosR.

It is also possible that the conformational flexibility of GAFB may be just an artefact from being an isolated domain, for example DosS GAFA as an isolated domain forms two species upon binding CO but only one conformation is formed in full length DosS (Ioanoviciu *et al.,* 2007). Alternatively DosS GAFB may play a role in dimerisation as in the case of GAFA PDE2A (Martinez *et al.,* 2002) and GAFA and GAFB of cyanobacterial adenylyl cylase (cynB2) (Martinez *et al.,* 2005). From the crystal structures of cynB2 and PDE2A shown in Figure 6-5 it can be seen that the GAFA/B dimer of cynB2 is antiparallel whereas the GAFA/B dimer of PDE2A is parallel, suggesting that there is more than one type of dimerisation interface for GAF domains. In both structures the tandem GAF domains are connected by a long helix which plays a role in dimerisation however the disposition of the two GAF domains within both structures differs. In PDE2A, GAFA forms the dimerisation domain and is far apart from the cGMP binding GAFB domain. The tandem domains are connected via a long helix, and within the dimer the C-terminal portion of this helix gradually bifurcates (Figure 6-5(A)). In contrast, within the

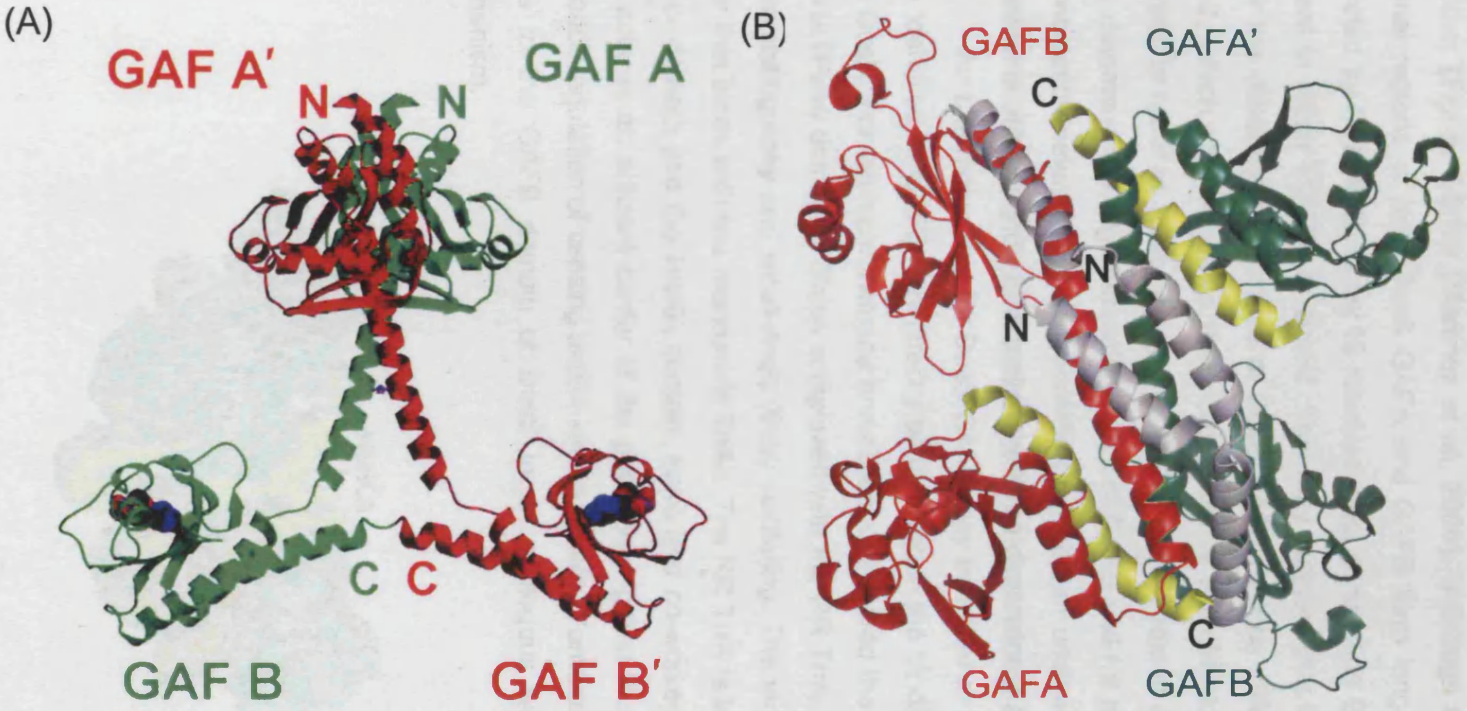


Figure 6-5: Dimerisation of GAF domains (A) PDE2A GAFA/B, cGMP is bound to GAFB. (B) cyaB2 GAFA/B, N-terminal helices are coloured grey, and the C-terminal helices coloured yellow. Figure (A) taken from Martinez et al., 2005.

cynB2 dimer both the N- and C- terminal helices run almost parallel to each other over their entire length. Dimerisation occurs mainly via the two N-terminal helices of GAFA allowing the GAFA domain of one monomer to pack tightly against the GAFB domain of the other (Figure 6-5(B)) (Martinez *et al.*, 2005). Although both the C- terminal and N-terminal regions of both DosS GAFA and GAFB form long α-helices, there is a linker predicted by PSIPRED, to be 13 residues long, connecting GAFA to GAFB, which in not present in either PDE2A or cynB2. This may suggest that if a dimerisation event does occur the dimerisation interface of DosS GAFB may be different from both PDE2A and cynB2, which also suggests that mechanism of histidine kinase activitation, would be different. In order to establish whether GAFB plays a role in dimerisation it is important to firstly determine the oligomerisation states of GAFB and full length DosS possibly through sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation (AUC) experiments and secondly to establish the three-dimensional structure of GAFB.

An alternative role for DosS GAFB may be either in binding a ligand that has not been identified or GAFB may directly be involved in the binding of the response regulator (RR) DosR. For example Yamada and colleagues solved the structure of the PAS, HisKA and HATPase domain of ThKA complexed with the RR TrrA, through a combination of X-ray crystallography and small-angle X-ray scattering. The structure comprises of a ThKA dimer that binds with two monomeric TrrAs. The RR TrrA is in direct contact with the PAS sensor domain and the HisKA domain. Shiro and co-workers proposed that the RR not only acts as an efficient carrier of the phosphoryl group, but also as an effector for the feedback regulation of sensing and/or kinase activity (Yamada *et al.*, 2006). Perhaps DosR binds to the GAFB domain of DosS in an analogous way to provide a feed back mechanism.

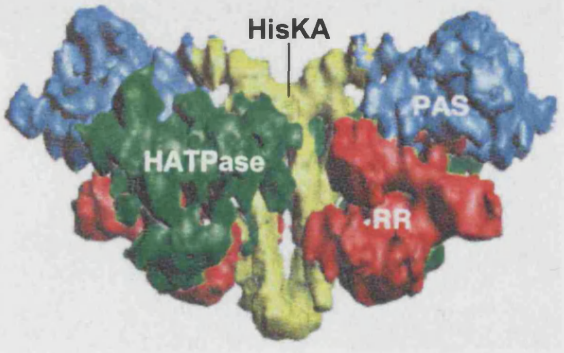


Figure 6-6: The Electron density map of the $(ThKA)_2$ / 2TrrA complex. ThKA: sky-blue-PAS domain; yellow-HisKA domain; green-HATPase domain. The RR TrrA is coloured red. Figure taken from Yamada el.al.,2006.

This structure also showed that the PAS domain comes into direct contact with the HATPase domain of the same subunit in the crystal structure, possibly indicating that ligand-induced structural changes of the PAS-sensor domain of one subunit are transmitted to the HATPase domain in the same subunit.

6.4 Concluding remarks

Mycobacterium tuberculosis bacillus are capable of remaining dormant, persisting within the human body for decades, within granulomas. The success of *MTBs* persistence during latency is thought to be due to its ability to adapt to an NO and hypoxic environment present within granulomas. The genetic response of *MTB* to hypoxia and NO is controlled by the sensor kinases DosS and DosT and the response regulator DosR that activates the expression of the DosR regulon; the precise molecular nature, to which, DosS responds was unknown prior to this work. The results presented within this thesis have shown that DosS is a haem-based sensor capable of binding both NO and oxygen. My findings provided a better understanding of how the DosS/DosR two component system functions and led to a proposed mechanism for histidine kinase activation. Hopefully, future work will build upon this data and lead towards the development of drugs to combat latent *MTB*.

CHAPTER SEVEN

REFERENCE LIST

7 Reference List

Agarwal,N., Woolwine,S.C., Tyagi,S., and Bishai,W.R. (2007). Characterization of the *Mycobacterium tuberculosis* sigma factor SigM by assessment of virulence and identification of SigM-dependent genes. Infect. Immun. *75*, 452-461.

Altendorf,K., Voelkner P, and Puppe W (1994). The sensor kinase KdpD and the response regulator KdpE control expression of the kdpFABC operon in *Escherichia coli*. Res Microbiol 374-381.

Aravind,L. and Ponting,C.P. (1997). The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem. Sci. *22*, 458-459.

Bacon,J., James,B.W., Wernisch,L., Williams,A., Morley,K.A., Hatch,G.J., Mangan,J.A., Hinds,J., Stoker,N.G., Butcher,P.D., and Marsh,P.D. (2004). The influence of reduced oxygen availability on pathogenicity and gene expression in *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) *84*, 205-217.

Baneyx,F. and Mujacic,M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. Nat. Biotechnol. *22*, 1399-1408.

Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L., Studholme,D.J., Yeats,C., and Eddy,S.R. (2004). The Pfam protein families database. Nucleic Acids Res. *32*, D138-D141.

Bertrand,T., Eady,N.A., Jones,J.N., Jesmin, Nagy,J.M., Jamart-Gregoire,B., Raven,E.L., and Brown,K.A. (2004). Crystal structure of *Mycobacterium tuberculosis* catalase-peroxidase. J. Biol. Chem. *279*, 38991-38999.

Betts,J.C., Lukey,P.T., Robb,L.C., McAdam,R.A., and Duncan,K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol. Microbiol *43*, 717-731.

Bradford,M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248-254.

Brennan,P.J. (2003). Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) *83*, 91-97.

Buck,M. (1998). Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. Q. Rev. Biophys. *31*, 297-355.

Calamita,H., Ko,C., Tyagi,S., Yoshimatsu,T., Morrison,N.E., and Bishai,W.R. (2005). The *Mycobacterium tuberculosis* SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. Cell Microbiol *7*, 233-244.

Campbell,I.A. and Bah-Sow,O. (2006). Pulmonary tuberculosis: diagnosis and treatment. BMJ *332*, 1194-1197.

Camus,J.C., Pryor,M.J., Medigue,C., and Cole,S.T. (2002). Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. Microbiology *148*, 2967-2973.

Cappelli,G., Volpe,E., Grassi,M., Liseo,B., Colizzi,V., and Mariani,F. (2006). Profiling of *Mycobacterium tuberculosis* gene expression during human macrophage infection: upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. Res. Microbiol. *157*, 445-455.

Cavanagh,J., Fairbrother,W.J., Palmer,A.G., and Skelton,N.J. (1996). Protein NMR Spectroscopy. (California.

Chan,J., Fan,X.D., Hunter,S.W., Brennan,P.J., and Bloom,B.R. (1991). Lipoarabinomannan, a possible virulence factor involved in persistence of *Mycobacterium tuberculosis* within macrophages. Infect. Immun. *59*, 1755-1761.

Charlet,D., Mostowy,S., Alexander,D., Sit,L., Wiker,H.G., and Behr,M.A. (2005). Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in *sigK*. Mol. Microbiol *56*, 1302-1313.

Clarkson,J. and Campbell,I.D. (2003). Studies of protein-ligand interactions by NMR. Biochem. Soc. Trans. *31*, 1006-1009.

Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry,C.E., III, Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N.,

Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,K., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,J., Squares,R., Squares,S., Sulston,J.E., Taylor,K., Whitehead,S., and Barrell,B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537-544.

Collins,D.M., Kawakami,R.P., de Lisle,G.W., Pascopella,L., Bloom,B.R., and Jacobs,W.R., Jr. (1995). Mutation of the principal sigma factor causes loss of virulence in a strain of the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. U. S. A *92*, 8036-8040.

Collins,H.L. and Kaufmann,S.H. (2001). The many faces of host responses to tuberculosis. Immunology *103*, 1-9.

Daffe,M. and Etienne,G. (1999). The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. Tuber. Lung Dis. *79*, 153-169.

Damberger,F., Nikonova,L., Horst,R., Peng,G., Leal,W.S., and Wuthrich,K. (2000). NMR characterization of a pH-dependent equilibrium between two folded solution conformations of the pheromone-binding protein from *Bombyx mori*. Protein Sci. *9*, 1038-1041.

Dasgupta,N., Kapur,V., Singh,K.K., Das,T.K., Sachdeva,S., Jyothisri,K., and Tyagi,J.S. (2000). Characterization of a two-component system, devR-devS, of *Mycobacterium tuberculosis*. Tuber. Lung Dis. *80*, 141-159.

De Carlo,S., Chen,B., Hoover,T.R., Kondrashkina,E., Nogales,E., and Nixon,B.T. (2006). The structural basis for regulated assembly and function of the transcriptional activator NtrC. Genes Dev. *20*, 1485-1495.

Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J., and Bax,A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277-293.

DeLano, W. L. The PyMOL Molecular Graphics System. 2002. San Carlos, CA, USA, DeLano Scientific.

DeMaio,J., Zhang,Y., Ko,C., Young,D.B., and Bishai,W.R. (1996a). A stationary-phase stress-response sigma factor from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A *93*, 2790-2794.

Dheda,K., Booth,H., Huggett,J.F., Johnson,M.A., Zumla,A., and Rook,G.A. (2005). Lung remodeling in pulmonary tuberculosis. J. Infect. Dis. *192*, 1201-1209.

Doukhan,L., Predich,M., Nair,G., Dussurget,O., Mandic-Mulec,I., Cole,S.T., Smith,D.R., and Smith,I. (1995). Genomic organization of the mycobacterial sigma gene cluster. Gene *165*, 67-70.

Dunham,C.M., Dioum,E.M., Tuckerman,J.R., Gonzalez,G., Scott,W.G., and Gilles-Gonzalez,M.A. (2003). A distal arginine in oxygen-sensing heme-PAS domains is essential to ligand binding, signal transduction, and structure. Biochemistry *42*, 7701-7708.

Dutta,R. and Inouye,M. (2000). GHKL, an emergent ATPase/kinase superfamily. Trends Biochem. Sci. *25*, 24-28.

Ericsson,U.B., Hallberg,B.M., Detitta,G.T., Dekker,N., and Nordlund,P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. *357*, 289-298.

Ewann,F., Jackson,M., Pethe,K., Cooper,A., Mielcarek,N., Ensergueix,D., Gicquel,B., Locht,C., and Supply,P. (2002). Transient requirement of the PrrA-PrrB two-component system for early intracellular multiplication of *Mycobacterium tuberculosis*. Infect. Immun. *70*, 2256-2263.

Ewann,F., Locht,C., and Supply,P. (2004). Intracellular autoregulation of the *Mycobacterium tuberculosis* PrrA response regulator. Microbiology *150*, 241-246.

Falquet,L., Pagni,M., Bucher,P., Hulo,N., Sigrist,C.J., Hofmann,K., and Bairoch,A. (2002). The PROSITE database, its status in 2002. Nucleic Acids Res *30*, 235-238.

Fenton,M.J. and Vermeulen,M.W. (1996). Immunopathology of tuberculosis: Roles of macrophages and monocytes. Infection and Immunity *64*, 683-690.

Fernandes,N.D., Wu,Q.L., Kong,D., Puyang,X., Garg,S., and Husson,R.N. (1999). A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. J. Bacteriol. *181*, 4266-4274.

Flynn,J.L. and Chan,J. (2001). Immunology of tuberculosis. Annu. Rev. Immunol. *19*, 93-129.

Galperin,M.Y., Nikolskaya,A.N., and Koonin,E.V. (2001). Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol. Lett. *203*, 11-21.

Gao,R., Mack,T.R., and Stock,A.M. (2007). Bacterial response regulators: versatile regulatory strategies from common domains. Trends Biochem. Sci. *32*, 225-234.

Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D., and Bairoch,A. (2005). Protein Identification and Analysis Tools on the ExPASY server. J.M.Walker, ed. Humana Press), pp. 571-607.

Gaynor,C.D., McCormack,F.X., Voelker,D.R., McGowan,S.E., and Schlesinger,L.S. (1995). Pulmonary surfactant protein A mediates enhanced phagocytosis of *Mycobacterium tuberculosis* by a direct interaction with human macrophages. J. Immunol. *155*, 5343-5351.

Gilles-Gonzalez,M.A., Ditta,G.S., and Helinski,D.R. (1991). A haemoprotein with kinase activity encoded by the oxygen sensor of Rhizobium meliloti. Nature *350*, 170-172.

Gilles-Gonzalez,M.A. and Gonzalez,G. (2005). Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. J. Inorg. Biochem. *99*, 1-22.

Gomez,J.E. and McKinney,J.D. (2004). *M. tuberculosis* persistence, latency, and drug tolerance. Tuberculosis (Edinb.) *84*, 29-44.

Gomez,M., Doukhan,L., Nair,G., and Smith,I. (1998). *sigA* is an essential gene in *Mycobacterium smegmatis*. Mol. Microbiol. *29*, 617-628.

Gonzalez,G., Dioum,E.M., Bertolucci,C.M., Tomita,T., Ikeda-Saito,M., Cheesman,M.R., Watmough,N.J., and Gilles-Gonzalez,M.A. (2002). Nature of the displaceable heme-axial residue in the EcDos protein, a heme-based sensor from *Escherichia coli*. Biochemistry *41*, 8414-8421.

Gonzalo,A.J., Maia,C., Ferrer,N.L., Barilone,N., Laval,F., Soto,C.Y., Winter,N., Daffe,M., Gicquel,B., Martin,C., and Jackson,M. (2006). The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in *Mycobacterium tuberculosis*. J. Biol. Chem. *281*, 1313-1316.

Graham,J.E. and Clark-Curtiss,J.E. (1999). Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). Proc. Natl. Acad. Sci. U. S. A *96*, 11554-11559.

Greenfield,N.J. (1996). Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. Analytical Biochemistry *235*, 1-10.

Grzesiek,S. and Bax,A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. J. Biomol. NMR *3*, 185-204.

Gupta,S. and Chatterji,D. (2005). Stress responses in mycobacteria. IUBMB. Life *57*, 149-159.

Gupta,S., Sinha,A., and Sarkar,D. (2006). Transcriptional autoregulation by *Mycobacterium tuberculosis* PhoP involves recognition of novel direct repeat sequences in the regulatory region of the promoter. FEBS Lett. *580*, 5328-5338.

Hahn,M.Y., Raman,S., Anaya,M., and Husson,R.N. (2005). The M*ycobacterium tuberculosis* extracytoplasmic-function sigma factor SigL regulates polyketide synthases and secreted or membrane proteins and is required for virulence. J. Bacteriol. *187*, 7062-7071.

Haydel,S.E., Benjamin,W.H., Jr., Dunlap,N.E., and Clark-Curtiss,J.E. (2002). Expression, autoregulation, and DNA binding properties of the *Mycobacterium tuberculosis* TrcR response regulator. J. Bacteriol. *184*, 2192-2203.

Haydel,S.E. and Clark-Curtiss,J.E. (2006). The *Mycobacterium tuberculosis* TrcR response regulator represses transcription of the intracellularly expressed Rv1057 gene, encoding a seven-bladed beta-propeller. J. Bacteriol. *188*, 150-159.

Haydel,S.E., Dunlap,N.E., and Benjamin,W.H., Jr. (1999). In vitro evidence of two-component system phosphorylation between the *Mycobacterium tuberculosis* TrcR/TrcS proteins. Microb. Pathog. *26*, 195-206.

He,H., Hovey,R., Kane,J., Singh,V., and Zahrt,T.C. (2006). MprAB is a stress-responsive two-component system that directly regulates expression of sigma factors SigB and SigE in *Mycobacterium tuberculosis*. J. Bacteriol. *188*, 2134-2143.

Henderson,R.A., Watkins,S.C., and Flynn,J.L. (1997). Activation of human dendritic cells following infection with *Mycobacterium tuberculosis*. J. Immunol. *159*, 635-643.

Himpens,S., Locht,C., and Supply,P. (2000). Molecular characterization of the mycobacterial SenX3-RegX3 two-component system: evidence for autoregulation. Microbiology *146 Pt 12*, 3091-3098.

Ho,Y.S., Burden,L.M., and Hurley,J.H. (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. EMBO J. *19*, 5288-5299.

Hu,Y. and Coates,A.R. (1999). Transcription of two sigma 70 homologue genes, *sigA* and *sigB*, in stationary-phase *Mycobacterium tuberculosis*. J. Bacteriol. *181*, 469-476.

Hu,Y. and Coates,A.R. (2001). Increased levels of *sigJ* mRNA in late stationary phase cultures of *Mycobacterium tuberculosis* detected by DNA array hybridisation. FEMS Microbiol. Lett. *202*, 59-65.

Hu,Y., Kendall,S., Stoker,N.G., and Coates,A.R. (2004). The *Mycobacterium tuberculosis sigJ* gene controls sensitivity of the bacterium to hydrogen peroxide. FEMS Microbiol. Lett. *237*, 415-423.

Hulko,M., Berndt,F., Gruber,M., Linder,J.U., Truffault,V., Schultz,A., Martin,J., Schultz,J.E., Lupas,A.N., and Coles,M. (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. Cell *126*, 929-940.

Inouye,M. (2006). Signaling by transmembrane proteins shifts gears. Cell *126*, 829-831.

Ioanoviciu,A., Yukl,E.T., Moenne-Loccoz,P., and Montellano,P.R. (2007). DevS, a Heme-Containing Two-Component Oxygen *Sensor of Mycobacterium tuberculosis*. Biochemistry *46*, 4250-4260.

Kanelis,V., Forman-Kay,J.D., and Kay,L.E. (2001). Multidimensional NMR methods for protein structure determination. IUBMB. Life *52*, 291-302.

Karakousis,P.C., Yoshimatsu,T., Lamichhane,G., Woolwine,S.C., Nuermberger,E.L., Grosset,J., and Bishai,W.R. (2004). Dormancy phenotype displayed by extracellular *Mycobacterium tuberculosis* within artificial granulomas in mice. J. Exp. Med. *200*, 647-657.

Karplus,K. and Hu,B. (2001). Evaluation of protein multiple alignments by SAM-T99 using the BAliBASE multiple alignment test set. Bioinformatics. *17*, 713-720.

Kelly,S.M., Jess,T.J., and Price,N.C. (2005). How to study proteins by circular dichroism. Biochim. Biophys. Acta *1751*, 119-139.

Kelly,S.M. and Price,N.C. (1997). The application of circular dichroism to studies of protein folding and unfolding. Biochim. Biophys. Acta *1338*, 161-185.

Kelly,S.M. and Price,N.C. (2000). The use of circular dichroism in the investigation of protein structure and function. Curr. Protein Pept. Sci. *1*, 349-384.

Kendall,S.L., Movahedzadeh,F., Rison,S.C., Wernisch,L., Parish,T., Duncan,K., Betts,J.C., and Stoker,N.G. (2004). The *Mycobacterium tuberculosis* dosRS two-component system is induced by multiple stresses. Tuberculosis (Edinb.) *84*, 247-255.

Khan,A.E. and Kimerling,M.E. (2006). Chemotherapy of tuberculosis. In Tuberculosis & Nontuberculous Mycobacterial Infections, M.Schlossberg, ed. (Philadelphia: McGraw-Hill), pp. 77-108.

Khorchid,A. and Ikura,M. (2006). Bacterial histidine kinase as signal sensor and transducer. Int. J. Biochem. Cell Biol. *38*, 307-312.

Krogh,A., Larsson,B., von Heijne,G., and Sonnhammer,E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567-580.

Mahler,H.R. and Cordes,E.H. (1969). Biological Oxidations. In Biological Chemistry, (New York: Harper International), pp. 554-621.

Malhotra,V., Sharma,D., Ramanathan,V.D., Shakila,H., Saini,D.K., Chakravorty,S., Das,T.K., Li,Q., Silver,R.F., Narayanan,P.R., and Tyagi,J.S. (2004). Disruption of response regulator gene, *devR*, leads to attenuation in virulence of *Mycobacterium tuberculosis*. FEMS Microbiol. Lett. *231*, 237-245.

Manganelli,R., Dubnau,E., Tyagi,S., Kramer,F.R., and Smith,I. (1999). Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. Mol. Microbiol. *31*, 715-724.

Manganelli,R., Voskuil,M.I., Schoolnik,G.K., Dubnau,E., Gomez,M., and Smith,I. (2002). Role of the extracytoplasmic-function sigma factor *sigma(H)* in *Mycobacterium tuberculosis* global gene expression. Mol. Microbiol. *45*, 365-374.

Manganelli,R., Voskuil,M.I., Schoolnik,G.K., and Smith,I. (2001). The *Mycobacterium tuberculosis* ECF sigma factor *sigmaE*: role in global gene expression and survival in macrophages. Mol. Microbiol. *41*, 423-437.

Marina,A., Waldburger,C.D., and Hendrickson,W.A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. EMBO J. *24*, 4247-4259.

Maris,A.E., Kaczor-Grzeskowiak,M., Ma,Z., Kopka,M.L., Gunsalus,R.P., and Dickerson,R.E. (2005). Primary and secondary modes of DNA recognition by the NarL two-component response regulator. Biochemistry *44*, 14538-14552.

Maris,A.E., Sawaya,M.R., Kaczor-Grzeskowiak,M., Jarvis,M.R., Bearson,S.M., Kopka,M.L., Schroder,I., Gunsalus,R.P., and Dickerson,R.E. (2002). Dimerization allows DNA target site recognition by the NarL response regulator. Nat. Struct. Biol. *9*, 771-778.

Martinez,S.E., Beavo,J.A., and Hol,W.G. (2002a). GAF Domains: Two-Billion-Year-Old Molecular Switches that Bind Cyclic Nucleotides. Mol. Interv. *2*, 317-323.

Martinez,S.E., Bruder,S., Schultz,A., Zheng,N., Schultz,J.E., Beavo,J.A., and Linder,J.U. (2005). Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: modes of ligand binding and dimerization. Proc. Natl. Acad. Sci. U. S. A *102*, 3082-3087.

Martinez,S.E., Wu,A.Y., Glavas,N.A., Tang,X.B., Turley,S., Hol,W.G., and Beavo,J.A. (2002b). The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. Proc. Natl. Acad. Sci. U. S. A *99*, 13260-13265.

Martinez-Hackert,E. and Stock,A.M. (1997). Structural relationships in the OmpR family of winged-helix transcription factors. J. Mol. Biol. *269*, 301-312.

Mascher,T., Helmann,J.D., and Unden,G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol. Biol. Rev. *70*, 910-938.

McGuffin,L.J., Bryson,K., and Jones,D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics. *16*, 404-405.

Michele,T.M., Ko,C., and Bishai,W.R. (1999). Exposure to antibiotics induces expression of the *Mycobacterium tuberculosis sigF* gene: implications for chemotherapy against mycobacterial persistors. Antimicrob. Agents Chemother. *43*, 218-225.

Miroux,B. and Walker,J.E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. *260*, 289-298.

Miyatake,H., Mukai,M., Park,S.Y., Adachi,S., Tamura,K., Nakamura,H., Nakamura,K., Tsuchiya,T., Iizuka,T., and Shiro,Y. (2000). Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J. Mol. Biol. *301*, 415-431.

Moreira,A.L. (2004). Bacille Calmette-Guerin: efficacy and immunity. In Tuberculosis, W.N.Rom and S.M.Garay, eds. (Philadelphia: Lippincott Williams & Williams), pp. 875-884.

Morth,J.P., Gosmann,S., Nowak,E., and Tucker,P.A. (2005). A novel two-component system found in *Mycobacterium tuberculosis*. FEBS Lett. *579*, 4145-4148.

Mukai,M., Nakamura,K., Nakamura,H., Iizuka,T., and Shiro,Y. (2000). Roles of Ile209 and Ile210 on the heme pocket structure and regulation of histidine kinase activity of oxygen sensor FixL from *Rhizobium meliloti*. Biochemistry *39*, 13810-13816.

Murzin,A.G., Brenner,S.E., Hubbard,T., and Chothia,C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536-540.

Muttucumaru,D.G., Roberts,G., Hinds,J., Stabler,R.A., and Parish,T. (2004). Gene expression profile of *Mycobacterium tuberculosis* in a non-replicating state. Tuberculosis (Edinb.) *84*, 239-246.

Neldon,D.L. and Cox,M.M. (2005). Oxidative Phosphorylation and Photophosphorylation. In Lehninger Principles of Biochemistry, W.H. Freeman and Company), pp. 690-750.

Nielsen,H., Engelbrecht,J., Brunak,S., and von Heijne,G. (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int. J. Neural Syst. *8*, 581-599.

Nowak,E., Panjikar,S., Konarev,P., Svergun,D.I., and Tucker,P.A. (2006a). The structural basis of signal transduction for the response regulator PrrA from *Mycobacterium tuberculosis*. J. Biol. Chem. *281*, 9659-9666.

Nowak,E., Panjikar,S., Morth,J.P., Jordanova,R., Svergun,D.I., and Tucker,P.A. (2006b). Structural and functional aspects of the sensor histidine kinase PrrB from *Mycobacterium tuberculosis*. Structure. *14*, 275-285.

Pace,C.N., Vajdos,F., Fee,L., Grimsley,G., and Gray,T. (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Sci. *4*, 2411-2423.

Parish,T., Smith,D.A., Kendall,S., Casali,N., Bancroft,G.J., and Stoker,N.G. (2003a). Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. Infect. Immun. *71*, 1134-1140.

Parish,T., Smith,D.A., Roberts,G., Betts,J., and Stoker,N.G. (2003b). The senX3-regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology *149*, 1423-1435.

Park,H. and Inouye,M. (1997). Mutational analysis of the linker region of EnvZ, an osmosensor in *Escherichia coli*. J. Bacteriol. *179*, 4382-4390.

Park,H.D., Guinn,K.M., Harrell,M.I., Liao,R., Voskuil,M.I., Tompa,M., Schoolnik,G.K., and Sherman,D.R. (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. Mol. Microbiol. *48*, 833-843.

Parkinson,J.S. and Kofoid,E.C. (1992). Communication modules in bacterial signaling proteins. Annu. Rev. Genet. *26*, 71-112.

Pearl,F., Todd,A., Sillitoe,I., Dibley,M., Redfern,O., Lewis,T., Bennett,C., Marsden,R., Grant,A., Lee,D., Akpor,A., Maibaum,M., Harrison,A., Dallman,T., Reeves,G., Diboun,I., Addou,S., Lise,S., Johnston,C., Sillero,A., Thornton,J., and Orengo,C. (2005). The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res. *33*, D247-D251.

Pelton,J.G., Kustu,S., and Wemmer,D.E. (1999). Solution structure of the DNA-binding domain of NtrC with three alanine substitutions. J. Mol. Biol. *292*, 1095-1110.

Perez,E., Samper,S., Bordas,Y., Guilhot,C., Gicquel,B., and Martin,C. (2001). An essential role for *phoP* in *Mycobacterium tuberculosis* virulence. Mol. Microbiol. *41*, 179-187.

Piddington,D.L., Kashkouli,A., and Buchmeier,N.A. (2000). Growth of *Mycobacterium tuberculosis* in a defined medium is very restricted by acid pH and Mg(2+) levels. Infect. Immun. *68*, 4518-4522.

Provencher,S.W. and Glockner,J. (1981). Estimation of globular protein secondary structure from circular dichroism. Biochemistry *20*, 33-37.

Pusey,M.L., Liu,Z.J., Tempel,W., Praissman,J., Lin,D., Wang,B.C., Gavira,J.A., and Ng,J.D. (2005). Life in the fast lane for protein crystallization and X-ray crystallography. Prog. Biophys. Mol. Biol. *88*, 359-386.

Rattle,H. (1995). An NMR PRIMER for Life Scientists. Partnership Press).

Rickman,L., Saldanha,J.W., Hunt,D.M., Hoar,D.N., Colston,M.J., Millar,J.B., and Buxton,R.S. (2004). A two-component signal transduction system with a PAS domain-containing sensor is required for virulence of *Mycobacterium tuberculosis* in mice. Biochem. Biophys. Res Commun. *314*, 259-267.

Roberts,D.M., Liao,R.P., Wisedchaisri,G., Hol,W.G., and Sherman,D.R. (2004). Two sensor kinases contribute to the hypoxic response of *Mycobacterium tuberculosis*. J. Biol. Chem. *279*, 23082-23087.

Robinson,V.L., Buckler,D.R., and Stock,A.M. (2000). A tale of two components: a novel kinase and a regulatory switch. Nat. Struct. Biol. *7*, 626-633.

Roccatano,D., Colombo,G., Fioroni,M., and Mark,A.E. (2002). Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study. Proc. Natl. Acad. Sci. U. S. A *99*, 12179-12184.

Rodgers,K.R. and Lukat-Rodgers,G.S. (2005). Insights into heme-based O2 sensing from structure-function relationships in the FixL proteins. J. Inorg. Biochem. *99*, 963-977.

Rodrigue,S., Provvedi,R., Jacques,P.E., Gaudreau,L., and Manganelli,R. (2006). The sigma factors of *Mycobacterium tuberculosis*. FEMS Microbiol Rev. *30*, 926-941.

Rybalkin,S.D., Rybalkina,I.G., Shimizu-Albergine,M., Tang,X.B., and Beavo,J.A. (2003). PDE5 is converted to an activated state upon cGMP binding to the GAF A domain. EMBO J. *22*, 469-478.

Saini,D.K., Malhotra,V., Dey,D., Pant,N., Das,T.K., and Tyagi,J.S. (2004a). DevR-DevS is a bona fide two-component system of *Mycobacterium tuberculosis* that is hypoxia-responsive in the absence of the DNA-binding domain of DevR. Microbiology *150*, 865-875.

Saini,D.K., Malhotra,V., and Tyagi,J.S. (2004b). Cross talk between DevS sensor kinase homologue, Rv2027c, and DevR response regulator of *Mycobacterium tuberculosis*. FEBS Lett. *565*, 75-80.

Saini,D.K., Pant,N., Das,T.K., and Tyagi,J.S. (2002). Cloning, overexpression, purification, and matrix-assisted refolding of DevS (Rv 3132c) histidine protein kinase of *Mycobacterium tuberculosis*. Protein Expr. Purif. *25*, 203-208.

Sasakura,Y., Hirata,S., Sugiyama,S., Suzuki,S., Taguchi,S., Watanabe,M., Matsui,T., Sagami,I., and Shimizu,T. (2002). Characterization of a direct oxygen sensor heme protein from *Escherichia coli*. Effects of the heme redox states and mutations at the heme-binding site on catalysis and structure. J. Biol. Chem. *277*, 23821-23827.

Schlesinger,L.S. (1993). Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. J. Immunol. *150*, 2920-2930.

Schlesinger,L.S., Bellinger-Kawahara,C.G., Payne,N.R., and Horwitz,M.A. (1990). Phagocytosis of *Mycobacterium tuberculosis* is mediated by human monocyte complement receptors and complement component C3. J. Immunol. *144*, 2771-2780.

Schultz,J., Milpetz,F., Bork,P., and Ponting,C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U. S. A *95*, 5857-5864.

Sekharan,M.R., Rajagopal,P., and Klevit,R.E. (2005). Backbone 1H, 13C, and 15N resonance assignment of the 46 kDa dimeric GAF A domain of phosphodiesterase 5. J. Biomol. NMR *33*, 75.

Sherman,D.R., Voskuil,M., Schnappinger,D., Liao,R., Harrell,M.I., and Schoolnik,G.K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. Proc. Natl. Acad. Sci. U. S. A *98*, 7534-7539.

Sopory,S., Balaji,S., Srinivasan,N., and Visweswariah,S.S. (2003). Modeling and mutational analysis of the GAF domain of the cGMP-binding, cGMP-specific phosphodiesterase, PDE5. FEBS Lett. *539*, 161-166.

Sousa,E.H., Tuckerman,J.R., Gonzalez,G., and Gilles-Gonzalez,M.A. (2007). DosT and DevS are oxygen-switched kinases in *Mycobacterium tuberculosis*. Protein Sci.

Spoerner,M., Nuehs,A., Ganser,P., Herrmann,C., Wittinghofer,A., and Kalbitzer,H.R. (2005). Conformational states of Ras complexed with the GTP analogue GppNHp or GppCH2p: implications for the interaction with effector proteins. Biochemistry *44*, 2225-2236.

Steyn,A.J., Collins,D.M., Hondalus,M.K., Jacobs,W.R., Jr., Kawakami,R.P., and Bloom,B.R. (2002). *Mycobacterium tuberculosis* WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. Proc. Natl. Acad. Sci. U. S. A *99*, 3147-3152.

Steyn,A.J., Joseph,J., and Bloom,B.R. (2003). Interaction of the sensor module of *Mycobacterium tuberculosis* H37Rv KdpD with members of the Lpr family. Mol. Microbiol. *47*, 1075-1089.

Stock,A.M., Robinson,V.L., and Goudreau,P.N. (2000). Two-component signal transduction. Annu. Rev. Biochem. *69*, 183-215.

Stock,J.B., Ninfa,A.J., and Stock,A.M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. *53*, 450-490.

Taylor,B.L. and Zhulin,I.B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol. Mol. Biol. Rev. *63*, 479-506.

Taylor,S.S., Knighton,D.R., Zheng,J., Ten Eyck,L.F., and Sowadski,J.M. (1992). Structural framework for the protein kinase family. Annu. Rev. Cell Biol. *8*, 429-462.

Thakur,K.G., Joshi,A.M., and Gopal,B. (2007). Structural and Biophysical Studies on Two Promoter Recognition Domains of the Extra-cytoplasmic Function {sigma} Factor {sigma}C from Mycobacterium tuberculosis. J. Biol. Chem. *282*, 4711-4718.

Thomas,M. (1997). Ultraviolet and Visible Spectroscopy. John Wiley & Son).

Thompson,J.D., Higgins,D.G., and Gibson,T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Tuckerman,J.R., Gonzalez,G., Dioum,E.M., and Gilles-Gonzalez,M.A. (2002). Ligand and oxidation-state specific regulation of the heme-based oxygen sensor FixL from *Sinorhizobium meliloti*. Biochemistry *41*, 6170-6177.

Tufariello,J.M., Chan,J., and Flynn,J.L. (2003). Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. Lancet Infect. Dis. *3*, 578-590.

Varnado,C.L. and Goodwin,D.C. (2004). System for the expression of recombinant hemoproteins in *Escherichia coli*. Protein Expr. Purif. *35*, 76-83.

Velasco-Velazquez,M.A., Barrera,D., Gonzalez-Arenas,A., Rosales,C., and Agramonte-Hevia,J. (2003). Macrophage *Mycobacterium tuberculosis* interactions: role of complement receptor 3. Microb. Pathog. *35*, 125-131.

Via,L.E., Curcic,R., Mudd,M.H., Dhandayuthapani,S., Ulmer,R.J., and Deretic,V. (1996). Elements of signal transduction in *Mycobacterium tuberculosis*: in vitro phosphorylation and in vivo expression of the response regulator MtrA. J. Bacteriol. *178*, 3314-3321.

Vidarsson,G. and van de Winkel,J.G. (1998). Fc receptor and complement receptor-mediated phagocytosis in host defence. Curr. Opin. Infect. Dis. *11*, 271-278.

Volz,K. (1993). Structural conservation in the CheY superfamily. Biochemistry *32*, 11741-11753.

Voskuil,M.I. (2004). *Mycobacterium tuberculosis* gene expression during environmental conditions associated with latency. Tuberculosis (Edinb.) *84*, 138-143.

Voskuil,M.I., Schnappinger,D., Visconti,K.C., Harrell,M.I., Dolganov,G.M., Sherman,D.R., and Schoolnik,G.K. (2003). Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J. Exp. Med. *198*, 705-713.

Voskuil,M.I., Visconti,K.C., and Schoolnik,G.K. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis (Edinb.) *84*, 218-227.

Walters,S.B., Dubnau,E., Kolesnikova,I., Laval,F., Daffe,M., and Smith,I. (2006). The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol. Microbiol. *60*, 312-330.

Wasserfallen,A., Huber,K., and Leisinger,T. (1995). Purification and structural characterization of a flavoprotein induced by iron limitation in Methanobacterium thermoautotrophicum Marburg. J. Bacteriol. *177*, 2436-2441.

Wayne,L.G. and Sohaskey,C.D. (2001). Nonreplicating persistence of *mycobacterium tuberculosis*. Annu. Rev. Microbiol. *55*, 139-163.

West,A.H. and Stock,A.M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem. Sci. *26*, 369-376.

Wisedchaisri,G., Wu,M., Rice,A.E., Roberts,D.M., Sherman,D.R., and Hol,W.G. (2005). Structures of *Mycobacterium tuberculosis* DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. J. Mol. Biol. *354*, 630-641.

Wishart,D.S., Bigam,C.G., Holm,A., Hodges,R.S., and Sykes,B.D. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J. Biomol. NMR *5*, 67-81.

Wishart,D.S. and Sykes,B.D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J. Biomol. NMR *4*, 171-180.

Wishart,D.S., Sykes,B.D., and Richards,F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry *31*, 1647-1651.

World Health Organisation. Who Report 2001: Global Tuberculosis Control. Technical Report. 2001.

Wu,Q.L., Kong,D., Lam,K., and Husson,R.N. (1997). A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. J. Bacteriol. *179*, 2922-2929.

Wu,S., Howard,S.T., Lakey,D.L., Kipnis,A., Samten,B., Safi,H., Gruppo,V., Wizel,B., Shams,H., Basaraba,R.J., Orme,I.M., and Barnes,P.F. (2004). The principal sigma factor *sigA* mediates enhanced growth of *Mycobacterium tuberculosis* in vivo. Mol. Microbiol. *51*, 1551-1562.

Yamada,S., Akiyama,S., Sugimoto,H., Kumita,H., Ito,K., Fujisawa,T., Nakamura,H., and Shiro,Y. (2006). The signaling pathway in histidine kinase and the response regulator complex revealed by X-ray crystallography and solution scattering. J. Mol. Biol. *362*, 123-139.

Zahrt,T.C. and Deretic,V. (2000). An essential two-component signal transduction system in *Mycobacterium tuberculosis*. J. Bacteriol. *182*, 3832-3838.

Zahrt,T.C. and Deretic,V. (2001). *Mycobacterium tuberculosis* signal transduction system required for persistent infections. Proc. Natl. Acad. Sci. U. S. A *98*, 12706-12711.

Zahrt,T.C., Wozniak,C., Jones,D., and Trevett,A. (2003). Functional analysis of the *Mycobacterium tuberculosis* MprAB two-component signal transduction system. Infect. Immun. *71*, 6962-6970.

Zoraghi,R., Corbin,J.D., and Francis,S.H. (2004). Properties and functions of GAF domains in cyclic nucleotide phosphodiesterases and other proteins. Mol. Pharmacol. *65*, 267-278.

Zuiderweg,E.R. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry *41*, 1-7.

CHAPTER EIGHT

APPENDIX

Table 8-1: Chemical shifts of ^{1}H, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, $^{1}H_{\alpha}$ resonances for $GAFB_{231-379}$ (ppm).

Residue		N	HN	CA	CB	CO	HA	HA2
1	P	-	-	-	-	-	-	
2	A	-	-	-	-	-	-	
3	T	-	-	-	-	-	-	
4	V	-	-	-	-	-	-	
5	F	-	-	-	-	-	-	
6	R	-	-	-	-	-	-	
7	L	-	-	-	-	-	-	
8	V	-	-	-	-	-	-	
9	A	-	-	-	-	-	-	
10	A	-	-	-	-	-	-	
11	E	-	-	-	-	-	-	
12	A	-	-	-	-	-	-	
13	L	-	-	-	-	-	-	
14	K	-	-	-	-	-	-	
15	L	-	-	-	-	-	-	
16	T	-	-	-	-	-	-	
17	A	-	-	-	-	-	-	
18	A	-	-	-	-	-	-	
19	D	-	-	53.9	43.7	175.9	4.77	
20	A	-	-	51.4	23.4	175.2	4.97	
21	A	-	-	51.5	23.1	174.9	5.63	
22	L	-	-	54.5	46.0	-	4.93	
23	V	-	-	58.7	33.8	183.5	-	
24	A	-	-	49.0	22.6	175.7	5.60	
25	V	-	-	57.7	33.8	-	5.03	
26	P	-	-	62.7	32.9	176.6	4.39	
27	V	119.9	8.10	64.9	32.2	175.8	3.73	
28	D	118.6	7.73	53.0	42.8	176.2	4.84	
29	E	123.2	8.73	57.6	30.4	176.2	4.15	
30	D	118.6	8.55	54.6	41.3	175.9	4.65	
31	M	121.3	7.46	53.6	34.2	-	4.70	
32	P	-	-	63.2	32.2	178.0	4.36	
33	A	127.0	8.46	55.6	19.4	179.3	4.05	
34	A	115.8	8.52	54.1	19.0	177.7	4.00	
35	D	115.6	7.90	54.2	42.6	176.1	4.70	
36	V	122.6	7.40	64.0	32.9	176.5	3.83	
37	G	114.4	8.88	45.5	-	175.0	-	
38	E	119.1	7.49	55.2	34.2	173.5	5.02	
39	L	119.3	8.82	52.1	45.9	174.2	4.57	
40	L	123.7	9.45	53.5	45.9	176.7	4.67	
41	V	126.4	8.83	63.1	30.7	178.7	4.33	
42	I	122.1	7.95	61.9	40.3	175.5	4.54	
43	E	119.9	7.19	56.8	33.4	173.4	4.55	
44	T	112.9	8.33	60.2	72.4	173.0	5.09	
45	V	119.0	8.98	62.6	36.1	174.4	4.33	
46	G	112.9	8.38	43.8	-	176.0	3.81	4.77
47	S	121.8	8.93	61.4	63.2	176.9	4.20	
48	A	122.2	8.72	54.3	19.3	178.6	4.21	
49	V	104.8	7.07	59.8	32.2	175.8	4.42	
50	A	125.1	7.27	55.2	19.1	179.8	4.18	
51	S	109.7	8.24	59.5	63.3	175.1	4.36	
52	I	117.7	7.48	61.3	37.9	175.0	3.95	
53	V	119.7	6.95	65.1	31.5	176.5	3.35	
54	G	115.6	8.86	45.0	-	174.1	3.57-	4.32
55	R	121.0	7.85	55.4	31.2	175.4	4.35	
56	T	114.4	8.10	59.9	70.8	174.6	5.32	
57	I	120.0	9.11	57.1	40.6	-	4.73	
58	P	-	-	-	-	-	-	
59	V	-	-	62.1	29.9	174.3	3.80	
60	A	127.3	7.31	51.9	21.5	177.4	4.50	
61	G	108.6	8.29	44.9	-	173.6	-	-
62	A	122.0	8.36	55.0	19.5	180.2	4.07	
63	V	117.9	8.09	66.4	31.8	177.2	3.76	
64	L	119.0	8.54	57.5	-	178.9	3.95	
65	R	116.7	7.73	60.8	-	178.1	3.84	

Table 8-1 *(continued)*

Residue		N	HN	CA	CB	CO	HA	HA2
66	E	119.6	7.82	60.1	29.7	178.5	4.06	
67	V	117.8	7.84	64.9	33.0	177.3	4.11	
68	F	115.5	8.36	60.4	40.6	174.1	4.18	
69	V	114.6	8.23	65.4	31.0	178.3	3.31	
70	N	113.1	7.28	53.4	39.6	176.4	4.83	
71	G	109.4	6.97	47.5	-	-	-	
72	I	123.2	7.93	57.4	39.9	173.8	4.50	
73	P	-	-	62.5	32.4	176.4	4.76	
74	R	120.0	8.58	53.4	35.5	173.6	4.76	
75	R	122.3	8.81	54.1	33.5	175.0	5.27	
76	V	116.0	8.78	58.6	35.8	176.2	4.93	
77	D	123.4	8.91	56.3	41.4	175.5	4.60	
78	R	116.2	7.39	54.8	32.7	-	-	
79	V	-	-	-	-	-	-	
80	D	-	-	52.8	40.9	175.1	4.70	
81	L	122.8	8.01	53.5	43.5	176.1	4.52	
82	E	126.2	8.45	58.0	29.8	176.6	4.04	
83	G	111.5	8.75	45.2	-	174.4	3.76	4.35
84	L	120.9	7.71	52.6	41.5	177.7	4.70	
85	D	123.1	8.66	57.6	-	-	4.24	
86	E	-	-	59.5	-	177.8	4.10	
87	L	116.5	7.82	54.3	41.5	177.2	4.37	
88	A	123.5	7.60	54.6	19.2	177.8	4.30	
89	D	115.5	8.05	53.9	41.2	176.1	4.77	
90	A	122.3	7.60	53.0	20.0	176.7	4.37	
91	G	107.4	7.97	44.6	-	-	-	-
92	P	-	-	63.1	34.7	174.9	4.40	
93	A	115.9	8.21	51.0	24.5	176.5	4.18	
94	L	116.0	8.45	53.1	43.8	175.4	4.67	
95	L	122.7	9.23	55.1	44.1	174.2	5.06	
96	L	121.1	9.17	50.1	-	-	5.04	
97	P	-	-	-	-	-	-	
98	L	-	-	-	-	-	-	
99	R	125.9	8.98	56.3	32.1	174.9	4.82	
100	A	122.2	8.28	52.1	21.7	176.7	4.45	
101	R	119.7	9.05	57.1	28.3	176.5	3.94	
102	G	107.0	8.69	45.8	-	174.1	3.68	4.12
103	T	115.3	7.82	59.7	71.3	173.2	4.67	
104	V	124.5	8.33	63.4	32.2	175.0	4.29	
105	A	129.1	9.22	51.9	21.3	176.9	4.62	
106	G	101.2	7.69	45.7	-	182.2	-	-
107	V	118.0	9.17	60.5	36.4	173.8	4.83	
108	V	125.7	9.25	60.6	-	174.5	4.95	
109	V	127.8	8.98	60.6	33.2	175.0	4.98	
110	V	127.3	9.24	61.5	33.7	174.3	4.57	
111	L	125.7	8.96	54.6	46.0	175.9	5.28	
112	S	115.0	8.38	56.8	65.7	174.4	-	
113	Q	125.4	9.66	56.6	29.3	177.1	4.54	
114	G	108.8	8.77	45.2	-	174.0	3.74	4.03
115	G	108.3	7.87	44.6	-	-	-	-
116	P	-	-	64.0	32.1	177.7	4.32	
117	G	109.4	8.50	45.3	-	174.1	3.85	
118	A	121.7	7.61	53.7	19.9	177.3	3.94	
119	F	112.8	7.62	55.8	43.3	176.2	5.54	
120	T	110.3	8.28	59.2	72.4	176.0	4.66	
121	D	121.8	9.18	57.7	40.0	178.5	4.43	
122	E	121.0	8.64	60.3	29.4	179.6	4.10	
123	Q	119.5	7.72	59.2	28.8	178.5	4.05	
124	L	120.0	7.83	58.9	41.7	177.3	3.88	
125	E	117.7	8.06	59.6	29.7	179.8	3.92	
126	M	119.6	8.20	59.1	33.3	179.7	4.21	
127	M	118.9	8.51	55.4	30.6	178.4	4.68	
128	A	124.6	8.90	55.2	18.6	-	4.26	
129	A	-	-	-	-	-	-	
130	F	-	-	-	-	-	-	

Table 8-1 *(continued)*

Residue		N	HN	CA	CB	CO	HA	HA2
131	A	-	-	-	-	-	-	
132	D	-	-	-	-	-	-	
133	Q	-	-	-	-	-	-	
134	A	-	-	-	-	-	-	
35	A	-	-	-	-	-	-	
136	L	-	-	-	-	-	-	
137	A	-	-	-	-	-	-	
138	W	-	-	-	-	-	-	
139	Q	-	-	-	-	-	-	
140	L	-	-	-	-	-	-	
141	A	-	-	-	-	-	-	
142	T	-	-	-	-	-	-	
143	S	-	-	-	-	-	-	
144	Q	-	-	-	-	-	-	
145	R	-	-	56.5	31.0	176.3	4.26	
146	R	121.8	8.21	56.3	31.5	176.3	4.32	
147	M	122.2	8.35	55.7	33.1	175.3	4.45	
148	R	126.9	7.86	57.6	31.8	-	4.14	

doi:10.1016/j.jmb.2005.09.011 J. Mol. Biol. (2005) 353, 929–936

JMB

Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

COMMUNICATION

A GAF Domain in the Hypoxia/NO-inducible *Mycobacterium tuberculosis* DosS Protein Binds Haem

Sunita Sardiwal[1], Sharon L. Kendall[2], Farahnaz Movahedzadeh[2] Stuart C. G. Rison[2], Neil G. Stoker[2] and Snezana Djordjevic[1]*